

Acting from Character:
How Virtue and Vice Explain Praise and Blame

by

Arden Ali
B.Phil., Philosophy and Politics
University of Pittsburgh, 2010

SUBMITTED TO THE DEPARTMENT OF LINGUISTICS AND PHILOSOPHY IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
SEPTEMBER 2016

© 2016 Massachusetts Institute of Technology. All rights reserved.

Signature redacted

Signature of Author:

Department of Linguistics and Philosophy
September 2, 2016

Signature redacted

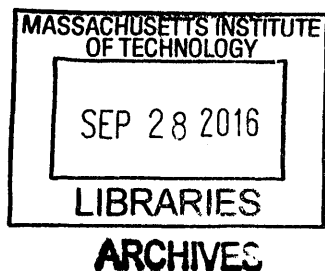
Certified by:

Kieran Setiya
Professor of Philosophy and Acting Chair
Thesis Supervisor

Signature redacted

Accepted by:

Roger White
Professor of Philosophy
Chair of the Committee on Graduate Students



Acting from Character
How Virtue and Vice Explain Praise and Blame
by Arden Ali

Submitted to the Department of Linguistics and Philosophy on September 2, 2016 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Philosophy.

ABSTRACT

This dissertation offers a theory of praise and blame: praiseworthy acts manifest virtue and blameworthy acts are incompatible with virtue. Despite its simplicity, proposals like mine have been largely ignored. After all, don't good people sometimes deserve blame, and bad people sometimes deserve praise? I believe the significance of this thought has been exaggerated. The chapters of this dissertation argue that we should understand praiseworthiness and blameworthiness by appeal to the concept of virtue, even granting the possibility of uncharacteristic behaviour.

Chapter One argues against the popular view of praiseworthiness, according to which acting well requires only that the agent is moved by the right reasons and acts rightly. At its most plausible, I claim, this view employs a concept of 'acting for the right reasons' that can only be understood in relation to virtue, e.g. someone acts for the right reasons just in case she is momentarily disposed as virtue requires, or has a disposition that approximates virtue. Praiseworthy acts are manifestations of virtue, perhaps qualified in some way, but nonetheless only intelligible in virtue-theoretic terms.

Chapter Two builds an account of blameworthiness. In response to puzzling cases of excuse, I distinguish *full* and *infallible* virtue. Roughly put: full virtue requires the disposition to act well; infallible virtue involves perfect compliance with the requirements of morality. This distinction allows us to articulate the relationship between character and culpability: blameworthy acts are those incompatible with full virtue in my sense.

Chapter Three addresses a conflict between my view and one dogma in the philosophy of responsibility. Philosophers usually distinguish mere badness and blameworthiness thusly: bad actions reflect deficiencies in one's ethical character but do not warrant resentment or indignation; blameworthy actions call for these attitudes. But I argue there is no privileged part of our psychology that can serve the role of 'ethical character' as it appears in the proposal. A better view falls out of the second chapter. On my view, there are two kinds of wrongdoing: those incompatible with full virtue, and those merely incompatible with infallible virtue. The former are blameworthy, but the latter are merely bad.

Thesis Supervisor: Kieran Setiya
Title: Professor of Philosophy and Acting Chair

Acknowledgments

For help in writing these chapters, I thank Nomy Arpaly, Nilanjan Das, Sanford Diehl, Ryan Doody, Sean Goedecke, Richard Holton, Brendan de Kenessey, Sofia Ortiz-Hinojosa, Bernhard Salow, Tamar Schapiro, Bradford Skow, Jack Spencer, Stephen Yablo, and audiences at the City University of New York and Australian Catholic University.

Special thanks are owed to my cohort at MIT and our counterparts at Harvard: Olivia Bailey, Dylan Bianchi, Jay Hodges, Douglas Kremm, Rebecca Millsop, Jack Marley-Payne, and Zeynep Soysal. I would not have survived graduate school without them. For help in philosophy and elsewhere, I am grateful also to Ekaterina Botchkina, Lyndal Grant, David Gray, Jillian Greis, Sophie Horowitz, Cole Leahy, Rose Lenehan, Matthew Mandelkern, Milo Phillips-Brown, and Ian Wells.

My committee deserves credit for the best parts of this work. Julia Markovits helped set this project in motion and has sustained it over the years. I thank her especially for her guidance in my early years of graduate school. I am fortunate to have met Sally Haslanger, who taught me a great deal about philosophy, and even more about leadership, fearlessness, and mentorship.

I would never have pursued philosophy if it were not for Kieran Setiya. This dissertation exists because of his encouragement and somewhat baffling confidence in my abilities. If I have any good qualities as a philosopher, I acquired them by imitating him.

To my family, near and far: thank you for the energy that each of you put into caring for me. To my mother in particular: thank you for your unshakeable and endless faith in me—you made my success possible in too many ways to count. Lastly, for seeing me through every hardship and happy moment, I owe my greatest and most unrepayable debt to Nashid Chaudhury. I feel lucky to have you in my life.

CHAPTER ONE

Manifestations of Virtue

Few philosophers think that the virtues bear moral value intrinsically.¹ Many believe that the moral value of a virtuous disposition is merely instrumental or derivative; it is morally good because it brings about the primary bearers of value, whatever those may be. Doubts about the role of virtue in moral philosophy are likely to follow: if the virtues are not themselves bearers of moral value, why should moral philosophers place any emphasis on them?

Occasionally, these doubts are part of *radical skepticism* about the intrinsic moral worth of all motivational or psychological states. Julia Driver (2001) calls this position ‘evaluational externalism.’ On her version of the view, the moral value of good motives or intentions rests solely in the outcomes that these states tend to produce.² For someone like this, it is no surprise that a virtuous disposition lacks any intrinsic value as well.

More often, however, skepticism about the virtues comes from more moderate voices. By and large, those who claim that the virtues are not intrinsically valuable are prepared to grant that good motives bear intrinsic value.³ They accept the intrinsic value of acting for the right reason but deny the intrinsic value of acting from virtue.

G.E. Moore expressed this *moderate skepticism*. For Moore, when someone acts from a

¹ Explicit doubts about the intrinsic moral value of the virtues appear in Moore (1903: 177), Sidgwick (1907: 393), Hooker (2002), Hurka (2006), Markovits (2010), and Crisp (2015). Crisp (2010: 23-4) says that a commitment to the intrinsic moral value of the virtues is the controversial claim of ‘explanatory virtue ethics.’

² Driver (2001:68). I discuss the view in §6.

³ My interest is *moral value* or *worth*. So henceforth when I claim that the virtues are ‘intrinsically valuable,’ this is shorthand for ‘the virtues bear *moral* value intrinsically.’

good motive, “it cannot be denied that the state of the man’s mind, in performing [the act], contains something intrinsically good” (1903: 177). But, in the same breath, Moore denounced Aristotle’s admiration for the virtues as a “gross absurdity” (1903: 176). He claimed that the exercise of a virtuous disposition “has, in general, no intrinsic value whatsoever” (1903: 176).

Moore’s view is alluring in part because it is only moderately skeptical. He denies that the virtues have intrinsic worth, but accepts that it is intrinsically valuable to act from good motives. Most modern-day skeptics echo him in this regard. Thomas Hurka, for example, describes a cowardly soldier who saves his comrades for the right reasons on one occasion.⁴ Hurka accepts that the soldier’s act is *morally better* because it is motivated by the right reasons, but denies that the act would have been *even better* had it issued from courage. He concludes that the moral value of a virtuous disposition is (mostly) instrumental since it tends “to produce individual virtuous acts, and feelings, and through them, further benefits such as pleasure for other people” (Hurka 2006: 73). Roger Crisp (2015: 14) describes the same moderate view. He imagines a malicious gangster, Ronnie, who once saves someone from danger for the right reasons. Although Crisp accepts the moral value of Ronnie’s good motives, he wonders: “...should we accept that Ronnie would have been *more* praiseworthy for acting in the way that he did had his action been based on a disposition? I think not” (2015: 14 his emphasis).

This chapter argues that moderate skepticism is *incoherent*. Any doubts about the moral value of acting from virtue must be part of radical skepticism. But since radical skepticism is itself untenable, there is little reason to doubt that the virtues bear moral value intrinsically.

Since moderate skepticism does not wear its incoherence on its face, my primary task is to bring out the confusion in the view. The position appears intelligible because the moderate

⁴ Hurka (2006: 72) and Hurka (2013: 12-13)

skeptic assumes that there are *two* ways of acting well on any occasion, namely acting for the right reason and acting from the relevant virtue. I argue that this assumption cannot be defended; there is only *one* way of acting well on a given occasion. As a result, it is impossible to maintain skepticism about the virtues without adopting skepticism about good motives too. It follows that there is no moderate post from which one can denounce the virtues. Our choice is to be radically skeptical or not skeptical at all.

My argument proceeds as follows. If there are two ways of acting well, then there is some sense in which acting from virtue differs from acting rightly for the right reason. In section one, I argue that this difference is not a difference *in kind*. In section two, I ask whether the two ways of acting well come apart *in degree*. To divorce the two ways of acting well, the skeptic relies on a principle about the relative breadth of the psychological dispositions involved in acting from virtue and acting for the right reason. But, in sections three, four, and five, I argue that the crucial principle is utterly unmotivated. In sections six and seven, I complete my case against skepticism about the virtues of character.⁵

1. Dispositions

Moderate skepticism is currently in vogue. This is partly because of recent work that draws a tight connection between acting well and acting for the right reason. Nomy Arpaly's insightful discussion put this connection on philosophy's collective radar.⁶ Julia Markovits draws the connection even more tightly than Arpaly; she claims that an act is morally worthy *just in case* it

⁵ My primary interest is skepticism about the moral worth of the virtues. I set aside the kind of skepticism that originates in situationist social psychology at this point, e.g. Harman (1999) and Doris (2002). I discuss Harman and Doris' kind of skepticism in §7.

⁶ Arpaly (2003: 67-116)

is performed for the right reasons.⁷ Versions of this claim give moderate skepticism its foothold. Granting that acting well requires acting for the right reason, a skeptic can ask: what value is *added* when someone acts from a virtuous disposition? This question has gone unanswered, so moderate skepticism has hardened in place. The skeptic seems to have shown that good motives are intrinsically valuable, but the virtues are not.

I am not going to answer the moderate skeptic's challenge to virtue theory directly. Instead, I argue that her challenge has no purchase on the virtue theorist. To show this, I target her assumption that there are two ways of acting well: acting rightly for the right reason and acting from the relevant virtue. Unless the skeptic assumes that these two ways of acting well come apart, her view makes no sense. Even though the view depends on this assumption, it has rarely been placed under careful scrutiny.⁸ The assumption warrants suspicion for two reasons.

First, the examples provided by Hurka and Crisp may be misleading. One might take them as proof that it is possible to act for the right reason without acting from virtue. If this possibility is real, the skeptic's assumption must be true. But the examples offered by Hurka and Crisp are too underdeveloped to establish this crucial possibility. Crisp, for example, just declares that Ronnie is "moved by real compassion, and genuinely wishes to help" despite failing to act "from a disposition or virtue" (Crisp 2015: 270). Without further elaboration, these examples do not establish the assumption that the skeptic needs.

Second, it is premature to accept the skeptic's assumption simply because acting from virtue involves a *trait* and acting for the right reason involves a *motivating reason*. It is not

⁷ See Markovits (2010). Compare Phillip Stratton-Lake (2000: 60), Korsgaard (2008:218-19), McGrath (2011: 130-7), and Sliwa (2015).

⁸ Brad Cokelet says that "any convincing defense of [the claim that acting from virtue is intrinsically valuable] must recognize the distinction between an agent's dispositions and [the agent's] occurrent motives..." (2015: 234). I agree that there is some difference between dispositional and occurrent states in general, but I will claim that this distinction doesn't do the work the skeptic needs.

obvious that these terms pick out different psychological kinds. Many traits appear to differ from other psychological dispositions merely in degree.⁹ Further, neither traits nor motivating reasons are irreducible psychological kinds. So, if acting from virtue and acting for the right reason involve different kinds of psychological states, there should be some description of this difference in terms of other states, like dispositions, beliefs, desires, and so on. Unless the skeptic can provide this description, her assumption should be regarded with suspicion.

Hurka suggests a description of precisely this sort. For him, the difference between acting from virtue and acting for the right reason is a difference *in kind*: an agent acts for the right reason just in case her act is caused by the appropriate *occurrent desire*, while an agent acts from virtue just in case she manifests a *disposition*.¹⁰ But this view can't be quite right. It seems to me that acting for the right reason must involve manifesting a disposition too. To see this, notice how quickly the view collapses in the light of the following two familiar difficulties.

First, if we treat acting for the right reason in Hurka's way, we cannot rule out causally deviant cases.¹¹ Earl is seated on a train and an elderly person boards. Earl notices that the elderly person needs the seat more than he does and he feels the desire to help someone in need. Earl's desire makes him feel good about himself, and while distracted with thoughts about his saintliness, he comes to think he missed his station and jumps up. The elderly passenger notices the seat and Earl offers it to him. Earl does the right thing, and his act was caused by (we may assume) the right desire, but even Hurka must agree that Earl does not act for the right reason. His desire does not bear the proper connection to his act.

Second, if we treat acting for the right reason in the way Hurka suggests, we cannot rule

⁹ See especially Allen & Potkay (1981), Fridhandler (1986), and Chapin (1988).

¹⁰ See Hurka (2006: 69-70)

¹¹ This case is modeled on cases in Davidson (1963: 78-79). For discussion of deviant causal chains, see Mele (2003:59-60).

out deviant patterns of reasoning. Suppose the fact that a joke will make you laugh is the right reason to tell it. Caligula might tell the joke because he wants to make you laugh. But Caligula wants to make you laugh because he knows that you recently had surgery on your abdomen and laughing will cause you tremendous pain. It is obviously a mistake to think Caligula tells the joke for the right reason. He has the appropriate desire, but his desire and action are connected by a deviant pattern of reasoning.¹²

In order to avoid cases like Earl and Caligula, Hurka must claim the following: one does not act for the right reason unless one's act is *caused in the right way* by the appropriate *noninstrumental* desire. The stipulation about the action's causal path rules out cases like Earl, while the stipulation about noninstrumentality rules out cases like Caligula. By this point, however, it seems that Hurka has abandoned the idea that the relevant psychological state is occurrent. These stipulations imply that agents who act for the right reason—like those who act from virtue—manifest dispositions. Here's why:

First, the expression of a noninstrumental desire requires the manifestation of a disposition. That is, an agent desires *p* noninstrumentally only if the agent manifests the appropriate disposition with respect to *p*. If I desire to ride my bicycle this evening for its own sake, then (for example) I must be disposed to ride my bicycle even if I have nowhere to go this evening. If I lack this disposition, it seems I don't desire to ride my bicycle for its own sake after all. Caligula desires to make you laugh, but he is *not* disposed to tell the joke if your abdomen is healthy and robust. This is precisely why his desire is merely instrumental. Someone who

¹² Notice that Hurka cannot avoid these cases by denying my supposition about the right reason to tell a joke. Cases like Caligula will exist even if Hurka identifies some other fact as the right reason to tell a joke. Suppose the fact that a joke will make you happy is the right reason to tell it. Caligula* may tell a joke because he wants to make you happy. But Caligula* wants to make you happy because he has covertly implanted a mechanism in your brain that will cause tremendous pain when it senses a suitable release of dopamine. Caligula* has the right occurrent desire, but he doesn't act for the right reason because his desire and action are connected by a deviant pattern of reasoning.

desires to make you laugh noninstrumentally is disposed to tell the joke even if laughing will cause you no pain at all.

Second, the stipulation about the action's causal path also seems to require abandoning Hurka's apparent proposal. Earl's gesture to the elderly passenger is caused by his desire to help someone in need, but the causal path between this desire and his action is deviant. To ensure that Earl doesn't act for the right reason, Hurka must insist that his desire causes his behaviour *in the right way*. But it seems to me that dispositions are again playing the crucial role. If one is thinking about the relevant kind of desires as dispositions, then Earl's desire and action are connected *in the right way* just in case his action is the manifestation of the disposition that constitutes his desire. Not every act caused by a disposition is a manifestation of that disposition. Consider an example from John Hyman: a man may knock himself unconscious by crashing his car, and he may crash his car because he is drowsy from having taken sleeping medication. In this case, Hyman notes, the man's unconsciousness is caused by the sleeping medication even though the medication's disposition to make him sleep is not manifested.¹³ Likewise, Earl's gesture is not the manifestation of a disposition to help someone in need, even though his action is caused indirectly by this disposition. Once again, Hurka must agree that acting for the right reason requires manifesting a disposition. This commitment is implicit in the reference to an act's being caused in the right way.

Hurka's simple way of identifying the difference between acting from virtue and acting for the right reasons does not work. Acting for the right reason, like acting from virtue, requires the manifestation of a disposition.¹⁴ So far, then, there is no difference in kind between the two.

¹³ See Hyman (2013: 12)

¹⁴ I am sympathetic to those who may prefer to identify the virtues with the states that ground a virtuous disposition and not the disposition itself. But my point here is simply that the psychological states involved in acting for the right reason *are the same kind* as the states involved in acting from virtue. If one decides that the states involved in

Perhaps acting for the right reason involves a different sort of disposition than acting from virtue. Some philosophers observe that virtues are ‘multi-track’ dispositions, or dispositions with more than one stimulus and/or manifestation condition. Ryle described hardness in this way: “when an object is described as hard, we do not mean only that it would resist deformation; we mean also that it would, for example, give out a sharp sound if struck, that it would cause pain if we came into sharp contact with it, that resilient objects would bounce off it, and so on indefinitely” (Ryle 1949: 44). Similarly, it may be that acting from virtue is multi-track; it requires dispositions to act, feel, perceive, think, detest, expect, and so on.¹⁵ Merely acting for the right reason, in contrast, may involve manifesting only one track among the many tracks of the virtuous person’s disposition.

I doubt that the moderate skeptic wants to distinguish acting from good character and acting for the right reason along these lines. Crisp grants that the one-time savior Ronnie is supposed to have thoughts and feelings that are indistinguishable from the virtuous person on the occasion that he acts (2015: 14). Ronnie is not merely disposed *to act* like the benevolent person, he is momentarily disposed *to feel and think* like her too. For Crisp, Ronnie fails to act from virtue even though he manifests all tracks of the virtuous person’s disposition.

There is a deeper problem here. If acting for the right reason involves manifesting one dispositional track among the many tracks that make up a virtue, then the dispositions that compose a virtue can be isolated from one another. But, if the dispositions that make up a virtue can be factorized in this way, then a virtue is not genuinely multi-track. A multi-track disposition is not a collection of discrete single-track dispositions, but a set of dispositions that cannot be

acting from virtue are beliefs and desires, then the states involved in acting for the right reason are also beliefs and desires. I choose here to say that the psychological states involved in both ways of acting well are dispositions, but nothing will ultimately hang on this decision.

¹⁵ See, for example, Attfield (1978) and Hursthouse (2013)

neatly separated from one another.¹⁶ Acting for the right reason cannot involve simply manifesting one dispositional track among the several that make up a virtue.

So far I have cleared up some distracting elements in the debate at hand. It is a mistake to rely on the occurrent/dispositional distinction or the multi-track/single-track distinction to capture the difference between people who act from virtue and those who merely act for the right reason. The difference between acting from virtue and acting for the right reason is not a difference in the *kind* of state from which the action issues.

2. Asymmetry

Is there a difference *in degree*? Is there some sense in which the agent who acts from virtue is *more* disposed than the one who merely acts for the right reason? Two pieces of terminology will keep things tidy:

An *R state* is the disposition manifested by someone who merely acts rightly for the right reason on a given occasion.

The respective *V state* is the disposition manifested by someone who acts from the relevant virtue on that occasion.

If someone can act for the right reason without acting from virtue, then it is possible to manifest an *R state* without manifesting the respective *V state*.

¹⁶ For discussion of multi-track dispositions and factorability, see Bird (2007: 21-24).

One striking point of agreement concerns the relative broadness of these states. Skeptics and non-skeptics assume that acting from good character requires a broader disposition than merely acting for the right reason.¹⁷ I call this assumption *Asymmetry*:

Asymmetry A *V* state is a broader dispositional state than the corresponding *R* states.

This is a promising attempt to divorce *R* states and *V* states in the way that moderate skepticism requires. But I don't think a skeptic about virtue can help herself to this principle. Before showing why, I should get clear on what the principle says.

One dispositional state is broader than another if the former manifests across more counterfactual situations than the latter.¹⁸ Both a champagne glass and a coffee mug are fragile. But the fragility of these two objects differs in an important respect. We might capture this difference by saying that the champagne glass has a *broader* disposition to break than the coffee mug.¹⁹ This is because the champagne glass manifests the disposition (i.e. it breaks) across more counterfactual situations than the coffee mug. That is, the glass breaks in all the cases where the mug does, and it also breaks in cases where the coffee mug will not, e.g. when dropped a short distance or filled with hot water, or sharply tapped with a spoon. This is how *Asymmetry* marks the difference between the person acting from good character and the person merely acting for

¹⁷ This principle is a more specific version of the general thought that traits are broad, or robust, dispositions for action. For discussion, especially regarding virtue ethics, see Harman (1999: 318), Doris (2002: 16–20), and Russell (2009: 246–252). See also Ross & Nisbett (1991: 59–89) and Pervin (1994:108).

¹⁸ For discussion of gradable dispositions, see Manley and Wasserman (2007).

¹⁹ One limitation of this example is that the set of possible worlds where champagne glasses break may not be a superset of the set of possible worlds where coffee mugs break. A more perfect example contrasts an *intact* coffee mug and a *slightly cracked* coffee mug. In this case, the cracked mug is more disposed to break when compared to the intact mug in the relevant sense; the cracked mug is disposed to break in all those cases where the intact mug does, and *more*.

the right reason. An agent who acts from good character is disposed to act rightly across *more* counterfactual situations than someone merely responding to the right reason.

Markovits (2010: 210) provides an example that brings out the intuitive force of *Asymmetry*. Imagine a fanatical dog-lover who saves a drowning stranger for the right reason, but who would ignore the stranger if his dog were present.²⁰ Suppose that the dog-lover manifests the *R* state but not the relevant *V* state. It must be uncontroversial that the dog-lover manifests a disposition to save the stranger in *some* counterfactual cases, where (for example) the stranger is wearing a different swimsuit, or screaming loudly, or has a different hair color. But, even if the dog-lover has a disposition to act rightly in those cases, *Asymmetry* explains that he can fall short of the *V* state. Being in the relevant *V* state requires having a disposition to save the stranger in all those cases *and more*. Someone who acts from courage is disposed to save the drowning stranger in all the cases that the dog-lover does, and others. A courageous person is disposed to save the drowning person even when her dog is present, or the water is choppy, or there's no one there to notice. So a proponent of *Asymmetry* claims.

It is crucial to see that *Asymmetry* has nothing to do with the temporal longevity of these two dispositions. More precisely, it does *not* say that *V* states are more enduring than *R* states. The endurance of the disposition is a poor way to divorce acting from good character and acting for the right reason.²¹ Of course, some *V* states are long-lasting, but others are short-lived. Imagine an agent who just reaches moral maturity and acts from courage on that occasion. Unfortunately, this newly courageous agent suffers serious trauma to his brain soon after his

²⁰ See Markovits (2010: 210)

²¹ It is fairly routine to couple broadness and temporal stability when talking about virtue. Hurka commits the dispositionalist to the claim that virtuous acts issue "from a *stable* disposition to act from similar motives *in similar circumstances*" (2006:71 my emphasis). Daniel Russell, for instance, identifies traits as "*temporally stable* dispositions to behave in certain ways *across a wide range of situations*" (Russell 2009: 268 my emphasis). Temporal stability and broadness in the sense of *Asymmetry* may correlate in some way, but it seems that they needn't appear together.

courageous act. As a result, he experiences a dramatic psychological shift, including loss of the disposition to act heroically. In this case, the *V* state is short-lived. In fact, this *V* state may be less enduring than its corresponding *R* states. Suppose, in that very case, an agent maintains the *R* state for a longer period of time than his ill-fated virtuous counterpart, perhaps simply because he avoids the brain damage. Everyone should agree that how long the disposition is in place doesn't distinguish *R* states and their respective *V* states. Indeed my opponents must think that some *R* states can be quite long-lasting; I may sustain a disposition to keep a promise I made years ago, and keep the promise on the relevant occasion, even though I am not disposed to keep any other promise. If there is a difference between *R* states and *V* states, stability over time is a poor way to identify it.

Many assume that endurance over time is the mark of a virtuous disposition, so I should dwell on this point a bit longer.²² Wouldn't it be strange to say that someone acts from virtue, but will never repeat the act or perform a similar act? As the case above illustrates, this must be possible. If I act from a virtue but subsequently lose the disposition because of trauma to my brain, I may never repeat any virtuous acts, but this would not cast doubt on whether I did act from virtue. Perhaps there is something strange about an agent who acts from virtue but loses her disposition *just a moment later*. If an agent's disposition fluctuates rapidly, it seems that she cannot act from virtue. But, crucially, this is also true of *R* states. If I am moved in one moment by my friend's pain, but lose my disposition in the next moment, one could reasonably doubt that I had the disposition in the first place. It may be *in general* rare for one's dispositional profile to

²² By denying that a virtuous disposition must endure over time, I depart from what Timpe & Boyd (2014: 6) call the 'traditional view' of virtue. Traditionalists who claim that virtues endure over time rarely say whether this is a conceptual claim about virtue or a psychological one. My argument targets the conceptual claim that virtues must be lasting dispositions. But, for someone like Aristotle, the virtues are very broad dispositions, so it may be *psychologically* unlikely that someone can manifest a virtue on one occasion and not shortly thereafter. I leave this psychological claim unscathed.

change radically in such a short period of time. So, even if one insists that the *V* state must show some stability, this minimal requirement will apply to all dispositions, including the *R* state.²³

The crucial difference between *V* states and *R* states has nothing to do with time. A claim about counterfactual broadness, i.e. *Asymmetry*, seems like the most promising alternative. The fate of moderate skepticism depends then on whether *Asymmetry* is true.

3. Against Asymmetry

Is *Asymmetry* true? This question cannot be answered independently of first-order moral theory. One cannot evaluate the relative broadness of *R* states and *V* states without adopting a substantive view about the content of these states. So the proper question is this: is *Asymmetry* true on any moral theory? In what follows, I argue that *Asymmetry* does not follow naturally from nearly any moral theory. To show this, I first show that *Asymmetry* is not entailed by a simple theory, and then I show that more complicated theories will have no interest in adopting the principle. Further, the single theory that has any sympathy for *Asymmetry* is utterly unsympathetic to skepticism about virtue. And so, the most promising way to pull *R* states and *V* states apart is either unmotivated or unavailable to the skeptic.

Recall that *Asymmetry* marks the difference between *R* states and *V* states just as we mark the difference between the fragility of a coffee mug and a champagne glass. Notice that the

²³ One might reply that actual stability over time isn't important, but possible stability over time is. *V* states are psychological states that *would* persist over time under normal conditions, while *R* states are not. Of course, there are some psychological states that are characteristically ephemeral, even under normal conditions. One might call these whims. Arpaly & Schroeder (2014) describe whims as "typically unstable, short-lived, and readily influenced by one's thoughts" (11). But *R* states are not like whims. Although one can offer some kind of explanation for an act performed on a whim, the explanation will not involve responsiveness to reasons. If I go shopping on a whim, and someone asks why I did it, the answer may be uninformative, e.g. 'I just did.' An agent who manifests the *R* state is not in a similar predicament. There must be an explanation of why he acted that shows his appreciation of the right-making fact. Compare some cases of arational action, e.g. Hursthouse (1991).

reason a coffee mug can be less fragile than a champagne glass is that *more than one* property can *make* something fragile. The properties that make a coffee mug fragile differ from the properties that make a champagne glass fragile, and so, the disposition to break may be broader in the latter case than in the former. Now consider a simple utilitarian account of what *makes* an action right. On this simple account, there is just *one* fact that makes an action right, namely the fact that the proposed action maximizes overall utility.²⁴ As a result, it seems to me, *R* states and *V* states must be equally broad on this theory. Here's why. Since there is only *one* right-making fact, the triggering conditions for the *R* state and *V* state must coincide. That is, since there are not two right-making facts on the theory, both dispositions must be triggered by the same (and only) right-making fact. Two dispositions that are triggered by the same fact will be equally broad. So, *R* states and *V* states are equally broad on a simple utilitarian theory. On this theory, then, *Asymmetry* is false.

Take the explanation of promise keeping on this simple utilitarian view: I ought to keep my promise because doing so maximizes overall utility. Suppose one asks: what disposition is required to keep the promise from the *V* state? Our imagined utilitarian only recognizes a single virtue: the disposition to maximize utility.²⁵ So, on this view, if I act from virtue in keeping my promise, then I manifest the disposition to keep the promise *when doing so maximizes utility*. Similarly, suppose one asks: what disposition is required to keep the promise from the *R* state? Again, there is only one right-making property, so it must be that someone acting from the *R* state is also disposed to keep the promise *when it maximizes overall utility*. The *R* state is exactly as broad as the *V* state. Since this result is not a peculiar feature of promise-keeping, the

²⁴ In the mug/champagne glass case, I referenced the existence of more than one fragile-making *property*. In the case of right-making, I talk instead about a single right-making *fact* instead of a single right-making *property*. This simply makes the prose easier to follow; nothing important hangs on this difference between the cases.

²⁵ On this point, see Foot (1985: 206).

reasoning runs equally well for any act. *Asymmetry* is false if this simple utilitarianism is true.

I just gave a quick argument that *Asymmetry* will not follow naturally from a very simple moral theory. There is one crucial point in the argument that the skeptic may resist:

In the argument above, I claimed that the simple utilitarian must say that the *R* state is the disposition to keep the promise when doing so maximizes overall utility. This is certainly a strong dispositional requirement on acting for the right reason. If an *R* state is a disposition to act in *every* case where keeping my promise maximizes utility, then I must be disposed to act well in very distant cases, including those when it is very difficult for me to keep the promise. One might deny that merely acting for the right reason requires such a broad disposition. More generally, one might deny that the *R* state is a disposition to act in *every* case where the right-making fact obtains.

If the skeptic resists my argument on these grounds, she faces a reasonable demand: if *R* states are not triggered in every case where the right-making fact obtains, *when is the disposition triggered?* I do not think she can give any satisfactory answer to this demand. It doesn't help to say that an *R* state is a disposition to act in the *nearby* possibilities where the right-making feature obtains. If *Asymmetry* is true, then it tells us which possibilities count as nearby; a nearby possibility is a case where an agent who merely acts for the right reason is disposed to act. But, since *Asymmetry* is precisely at issue, an alternative account can't just rest on nearness in this sense. What other answer can she give to my demand?

Instead of answering the demand, I suspect that a skeptic will try to deny that my demand gets a grip on her proposal in the first place. The demand to identify exactly which possibilities trigger the *R* state is misguided since (in her view) the breadth of the disposition is largely irrelevant to whether someone acts for the right reason. The important requirement is not

dispositional, but cognitive: in order to act for the right reason, one must merely *take* the right-making fact as the ground for her behavior. Let me spell out this response:

Not every fact that motivates someone counts among the things that the agent takes to explain what he is doing. The fact that I take to explain my swallowing an aspirin may be that I have a headache, even though there are other facts that play some role in motivating me, e.g. the fact that I am not allergic to aspirin, or not taking a contraindicated medication.²⁶ These facts are not the ones that I take to explain what I am doing. The ground for my behavior is that I have a headache. On the replacement under consideration, the *R* state may be a narrower disposition than the *V* state, but the agent who acts for the right reason must *take* the right-making fact as her reason in this privileged sense. In other words, the right-making fact must be the fact that she takes to explain what she is doing.²⁷ So, the view relaxes the dispositional requirement by insisting that the agent meet a cognitive one.

But, this proposal is plainly too permissive. That is, a cognitive requirement cannot compensate for a relaxed dispositional one.²⁸ Suppose Mia is disposed to keep her promise when

²⁶ One notable class of these kinds of facts are what Jonathan Dancy calls ‘enablers.’ See Dancy (1993: 22–26), (2004: 38–41).

²⁷ Some think that intentional action is performed under ‘the guise of the good.’ If so, when someone acts for a reason, she has a belief about what *justifies* her act; that is, she takes the right-making fact *as* right-making (e.g. Korsgaard 2008: 213–14). Here I merely claim that the agent’s belief is about what *explains* her behavior, rather than what justifies it (Setiya 2007: esp. 59–61; 2010). The worry I raise for the cognitive alternative to my strong dispositional requirement will stand even if one adopts the guise of the good claim. The general worry I raise is that the cognitive alternative allows agents who merely seem to have tricked or misled themselves into believing they are acting for the right reason count as truly acting for the right reason. This criticism extends to the ‘guise of the good’ version of the cognitive alternative too. On the guise of the good version, the agent also recognizes her reason as right-making. But presumably someone can believe that a reason is right-making, and believe they are acting for that reason, yet be entirely delusional. If the agent’s is in fact dispositionally sensitive to some other fact altogether, it doesn’t seem to matter whether she recognizes the right-making *as* right-making. She still does not act for the right reasons.

²⁸ To be clear, I am only advancing this strong dispositional requirement as a requirement on acting for the *right* reasons. It is a further question whether the dispositional requirement holds for acting for reasons in general. Though I do not argue for it here, this extension does seem *prima facie* plausible. It seems to me that someone who goes to the gym in order to improve her health must be disposed to go to the gym when it will improve her health (cf. Arpaly 2003: 88). As I clarify on p.15-16, this is not the same as requiring that she *would* go to the gym in all cases where it improves her health.

it is in her selfish interest to do so. In some cases, keeping the promise serves both her selfish interest *and* maximizes utility. Suppose that Mia keeps her promise on one of these occasions and believes that she is acting solely in order to maximize utility. In that sense, she *takes* the right-making fact as the ground for her behavior. On the alternative just proposed, Mia keeps her promise for the right reason. But, intuitively, Mia is simply *deluded* about what is motivating her. It seems to me that she is either ignoring her selfishness or utterly unreflective about it. If the proposal counts her among those who act for the right reason, it should be rejected.²⁹ And, if this proposal fails, my demand that the skeptic identify which possibilities trigger the *R* state remains. I do not see any other way that the skeptic can meet this demand, so she doesn't truly present an objection to the argument at the beginning of this section.

Still, my claims may appear suspicious. Doesn't it make it *too hard* to act for the right reason? If simple utilitarianism is true, anyone who acts for the right reason must manifest a disposition to so act *whenever* doing so maximizes utility. I gather that part of the uneasiness here has nothing to do with my claim about *R* states, but really targets what utilitarianism identifies as the right-making fact. As others have noted, utilitarianism faces the problem that the facts it says justify our actions shouldn't be the facts that motivate us.³⁰

Any residual concern rests on cases like Markovits' fanatical dog-lover. She claims that the dog-lover acts for the right reason, but she stipulates that he *would not* save the drowning

²⁹ One could respond that Mia's sort of delusion is impossible. Perhaps she cannot genuinely believe she is acting in order to maximize utility while being disposed to keep the promise only when it serves her interest. It may be that agents cannot form beliefs about why they are acting that are so drastically disoriented with respect to their actual dispositions. But, granting that Mia is an extreme example, it seems the alternative view is in trouble as long as even mildly delusional cases are possible. Imagine someone who is disposed to give gifts only when they expect one in return. Since gifts tend to make her friends happy, and they also reliably tend to return the favour, it is not difficult to imagine that this agent can trick herself into believing she is acting in order to bring joy to her friends. But it seems like a mistake to conclude she acts for that motive. So, unless even mild cases like these are also impossible, the alternative view will turn out to be too permissive.

³⁰ In particular, see Stocker (1976).

person if his dog were present.³¹ Isn't the dog-lover a counterexample to my claim about the breadth of *R* states? On my account, it needn't be true that the dog-lover *would* save the drowning person if his dog were present. My account says that the dog-lover must be *disposed* to save the person if his dog were present, or else he isn't acting for the right reason. To say that he is disposed in this way is not the same as saying that he *would* save the person in this case.³² After all, the connection between dispositions and counterfactual conditionals is notoriously messy. A glass is still disposed to break when dropped even though it *would not* break when dropped in some cases, like when it is covered in bubble wrap.³³ An agent may be disposed to act rightly when the right-making fact obtains even though he *would not* act rightly in some cases where the fact obtains. So the dog-lover is not a straightforward problem for the view.

Markovits may insist that the dog-lover can act for the right reasons even though he wouldn't even *be moved* to save the drowning person if his dog were present. But I think there is good reason to disagree. If one relaxes my dispositional requirement on acting for the right reason in order to make space for the dog-lover, there will be no way to rule out obvious cases of misaligned values. Consider, for instance, a religious fanatic who believes that non-believers deserve death, but believers do not.³⁴ One day, this fanatic saves a stranger, but would not have been moved to save him if the stranger were not wearing orthodox religious clothing. I take it that the religious fanatic is misguided about what matters morally and so does not act for the

³¹ One objection to my account of *R* states may be that very few people will act for the right reason. This is certainly true on simple utilitarianism. But, if moral theory makes room for more virtues (as many do), it will become less demanding to act for any particular right reason. If anything, the objection gives us good reason to prefer moral theories that provide more fine-grained virtues over those who provide only broad ones. See §7 for more on this point.

³² I am not saying that these counterfactuals have no relevance whatsoever. They may be relevant because they loosely track how the agent is disposed.

³³ For masking cases of this sort, and discussion of dispositions and counterfactuals, see Johnston (1992), Martin (1994), Lewis (1997), Bird (1998), Fara (2005), and Manley & Wasserman (2008; 2012),

³⁴ Compare Arpaly's 'extremist' in (2003: 74).

right reason. Yet, in order to accommodate the dog-lover who is not even moved to act when his dog is present, Markovits must admit that the religious fanatic acts for the right reason. But, as she must agree, this is the wrong result.

So far I have argued that *Asymmetry* does not follow from simple utilitarianism and defended the crucial claim about *R* states. Does *Asymmetry* fare better on a different moral theory? In the next section, I take up this question.

4. Beyond Utilitarianism

Simple utilitarianism identifies only a single property as right-making. A more complicated theory might claim that right-making properties may be *more* or *less* fundamental. Does *Asymmetry* follow from a view like this?

To make things concrete, imagine a contractualist explanation of my obligation to visit a sick friend.³⁵ Like simple utilitarianism, contractualism is a version of monism; it identifies a single fundamental right-making property. For a contractualist, the *most* fundamental reason to visit my sick friend is that I have no reasonable grounds on which I could justify failing to visit him. Unlike simple utilitarianism, however, a contractualist may identify *less* fundamental right-making reasons. He may point out that there are reasons in virtue of which I cannot justify failing to visit, e.g. I share a relationship with my friend, or I know how to make him laugh, or there is no one else at his bedside. These facts are right-making reasons, albeit less fundamental ones. One way to save *Asymmetry* is to claim that the *V* state is responsive to *more* fundamental right reasons, while the *R* state is responsive to *less* fundamental right reasons. But what could

³⁵ The view I have in mind is developed in Scanlon (1998).

motivate this move?

One might insist that less fundamental facts are sometimes right-making reasons in order to avoid charges of *alienation*. An agent who visits because he cannot justify doing otherwise is alienated from what is morally valuable in the case.³⁶ But if this is the motivation for treating *R* states in this way, the reasoning seems to extend to *V* states as well. After all, someone who acts from virtue is not alienated from what is morally valuable in the case *either*. So, worries about alienation may give us reason to treat both *R* states and *V* states as narrow dispositions, but it will not give us reason to adopt *Asymmetry*.

Another reason to think that *R* states respond to less fundamental reasons concerns the role of theory in everyday moral practice. It must be possible to act for the right reason without knowing anything about a moral theory like contractualism. This is possible only if *R* states are responsive to less fundamental reasons. But again, the reasoning extends to *V* states as well.³⁷ It is just as plausible that one can act from virtue without learning about moral theory. In fact, one might think moral theory is antithetical to the proper exercise of a virtue.³⁸ So, concern about the role of theory in everyday practice may motivate changes to the *R* state *and* *V* state, but it will not yield an asymmetric relationship between the two.

These are not the only considerations that might motivate this imagined contractualist to treat *R* states in the way she does. But they are representative of the reasons that might motivate

³⁶ For discussion of this point, see Stocker (1976), Baron (1984), Railton (1984), and Smith (1994: 71).

³⁷ Julia Annas (2011: 29) and Alison Hills (2015) argue that a virtuous person can *articulate* her reasons for acting. Notice that it is one thing to require that the agent is capable of articulating her reasons, and quite another to say that she must respond to a more *fundamental* reason. In fact, anyone who places an articulation condition on virtue will not want to treat *V* states as responsive to more fundamental facts, since this would make an articulation condition quite implausible. It may be plausible that virtuous people can say *something* about why they do the right thing, but it is unlikely that they can identify the morally fundamental reasons for acting as they do.

³⁸ Murdoch (1970), for instance, rejects the idea that virtue issues from an “impersonal quasi-scientific knowledge” but rather from “a refined and honest perception of what is really the case, [and] a patient and just discernment and exploration of what confronts one” (37).

someone in her direction. Neither yield an asymmetric relationship between *R* states and the respective *V* state. One is hard-pressed to find a reason to treat *R* states and *V* states differently. So, again, there is no easy way to get from the moral theory to *Asymmetry*.

How does *Asymmetry* fare once we abandon monism? One might think that *Asymmetry* will follow if there is more than one fundamental right-making property. On Ross' influential theory, there are several fundamental yet unconnected right-making reasons. He thought that the right reason to keep my promise is simply that I promised. On my way of speaking, the *R* state is the disposition to keep my promise when I have genuinely promised (1930: 38). The *V* state in this case is called *fidelity*. The *V* state just is "the disposition to fulfil promises.... [when] we have made them" (1930: 22).³⁹ The *mere number* of fundamental reasons does not help to establish an asymmetry in the broadness of *V* states and corresponding *R* states. The number of fundamental reasons just identifies how *many* distinct *R* states and *V* states there are. It does not tell us that one state is broader than the other. Yet again, there is no clear route from pluralism to *Asymmetry*. Pluralists about the structure of morality have no reason to favour pulling *V* states and *R* states apart.

5. Particularism

What are the prospects for *Asymmetry* if particularism is true? A particularist thinks that moral reasons do not hang together in any principled way.⁴⁰ Unsurprisingly, there isn't an obvious route to *Asymmetry* from this view. Of course, a particularist may claim that there are many

³⁹ Ross uses 'because' instead of 'when' but he draws attention to the fact that fidelity is a disposition to keep one's promises "irrespective of motive" (1930: 22).

⁴⁰ Particularism has been developed in Dancy (1993) and Dancy (2004). See also the essays included in Hooker and Little (2000).

cross-cutting and unrelated reasons to act on any given occasion. A broader disposition will be required to respond to some of these reasons and a narrower one will do for others. But, even so, there is no way to get from this conclusion to *Asymmetry*. By the particularist's own lights, these reasons bear no relation to one another, so we cannot identify *V* states as broader than *respective R* states. The upshot of the particularist's claim is that there are *many* virtues, some of which are highly specific. The purported asymmetric relationship between *R* states and *V* states remains unfounded.

At this point, things look dire for *Asymmetry*. To complete my case against it, I want to turn to the following question: how would a moral theory have to look in order for the principle to follow naturally? There is a special kind of particularism that yields the conclusion that *R* states are narrower than respective *V* states. But I am going to argue that *Asymmetry* comes at the cost of giving up moderate skepticism.

Consider a particularist who takes the idea of *sensitivity* seriously. For this particularist, each virtue is a capacity that allows its possessor to grasp an entire class of morally relevant facts.⁴¹ Each of these facts is a right-making reason because it is discernable by someone with the relevant virtue. Since the *V* state is a general capacity, one can assume that the virtuous person has a broad disposition to act. But, since it may be possible to act for one of the right-making reasons without exercising the virtue, the *R* state may be a narrower than the respective *V* state.

I grant that an asymmetry between *R* states and *V* states follows from this view. But, although the capacity view of virtue generates this asymmetry, it seems that the view cannot survive if it *also* denies that the virtues are the primary bearers of moral value. Here's why. The

⁴¹ The kind of view described here appears in McDowell (1979).

problem for the earlier version of particularism is that it couldn't match *R* states and *V* states together, and since *Asymmetry* is a claim about *respective R* and *V* states, there was no space for the principle. The capacity version of particularism gets around this concern by matching *R* states and *V* states conceptually: an *R* state is a disposition to respond to a particular fact, and this fact is right-making because it is discernable as such by someone with the relevant *V* state.⁴² *V* states and *R* states match up because of this relationship. A right-making reason *just is* a fact that is discernable by someone with the respective virtue.

But, if the *V* state is conceptually connected to the *R* state in this way, any grounds for moderate skepticism have crumbled. A moderate skeptic grants that *R* states bear moral value intrinsically, but denies that *V* states do. The kind of particularism that yields *Asymmetry* cannot say this.⁴³ According to the view, the *R* state would have no significance if it did not bear a relationship to the relevant *V* state. If this is so, the moral value of the *R* state is derived from that of its corresponding *V* state. When the particularist claims that the *R* states and *V* states are conceptually connected, she makes room for *Asymmetry*, but she gives up on skepticism about the intrinsic value of *V* states.

An analogy may help. I am interested in what bears value from the *moral* point of view. But one might ask a similar question about what bears value from a *practical* point of view. Compare a surgeon and her novice student.⁴⁴ Both the surgeon and the student may apply pressure to a wound in order to stop the patient's bleeding. Both may be disposed to apply pressure when the patient is bleeding in this way. But the trained surgeon's disposition differs

⁴² Notice that the conceptual relationship could not run in reverse. If *V* states have moral status because of their relationship to *R* states, the particularist gives up on her insistence that the moral realm is *unprincipled*.

⁴³ And this particularist may have no interest in doing so. McDowell develops the view as an alternative to the tradition that gives virtue merely a secondary role in moral theory (1979: 331).

⁴⁴ This analogy depends on understanding virtues as skills (see Jacobson 2005; Annas 2011). I take it that this analogy is appealing to the particularist because skills combine the perceptual, motivational and cognitive elements that they often use to describe the capacity involved with virtue.

from that of her student. When the surgeon applies pressure, she is of course sensitive to the bleeding, but she is also responding to the patient's blood pressure, the location of the wound, and whether such pressure might cause further damage to the patient. The student who applies pressure in order to stop the bleeding is not sensitive to this class of facts because she lacks the capacity that the surgeon has.

The capacity view of virtue pulls the *R* state and *V* state apart in a similar way. The *R* state involves responding to a right-making fact, but the *V* state requires sensitivity to a class of facts that are relevant to the case at hand. In the analogy, it is obvious that the primary bearer of practical value is the surgeon's skill, not the student's approximation of it. The practical value of the surgeon's capacity is not that it merely *brings about* acts like those performed by the student. Rather the reverse is true: we praise the student's practical acumen *because* she is acting like the surgeon she hopes to become.⁴⁵ Likewise for the capacity view of virtue; the value of the *R* state is derived from that of the *V* state.

Let's take stock. I claimed that the most promising way to pull *R* states and their respective *V* states apart is *Asymmetry*, a claim about the relative broadness of these states. But *Asymmetry* does not follow naturally from nearly any moral theory. The kind of particularism that says *R* states and *V* states may diverge will rule out any grounds for moderate skepticism in any case. It appears that the skeptic's crucial principle is not available to her.

6. Skepticism about Virtue

At the outset of this chapter, I identified three possible views about the moral value of the virtues

⁴⁵ Of course, we might also praise the student for acting well, *morally speaking*, in addition to praising her for acting well, *practically speaking*.

of character. For the radical skeptic, no internal state bears moral value intrinsically and so the virtues too have no intrinsic moral worth. For the moderate skeptic, good motives have intrinsic moral worth, but the virtues do not. And for the minority of non-skeptics, acting from a virtue has intrinsic moral value.

This debate is not three-pronged, but two-pronged. The moderate position rests on the assumption that there are two ways of acting well on a given occasion. The skeptic's best attempt to divorce these two ways of acting well is a principle I called *Asymmetry*. But I showed that simple utilitarianism did not have the resources to generate *Asymmetry* and the principle could not be motivated on other theories without giving up on skepticism. Once this is recognized, the incoherence of the moderate position is obvious: if the two states do not come apart, they must either both have value or else neither has value. Our choice is between *radical skepticism* and *no skepticism*.

Is radical skepticism plausible? Julia Driver shows some sympathy for the radical position. She defines an 'evaluational externalist' as someone who thinks that "the moral quality of a person's action or character is determined by factors external to agency, such as actual (rather than expected) consequences" (2001: 68). She says this position places "no intrinsic value on states of mind, or on factors that are internal to agency" (2001: 68).

I am not interested in refuting radical skepticism. But, after abandoning the moderate view, some skeptics may be prepared to take up the more radical position. This retreat would be ill-advised. The main feature that made moderate skepticism attractive is that it did not face the same difficulties as the radical position. To start, the radical view must explain the intuition that agents who act rightly for the right reason are *more praiseworthy* than agents who act rightly for

the wrong reason. This intuition strongly suggests that motives have intrinsic moral value.⁴⁶

How can a radical skeptic reply?

She may appeal to extrinsic moral value.⁴⁷ A motive, she may argue, can have value grounded in its extrinsic or relational properties, like (for example) being such that an agent with the motive is likely to bring about good outcomes.⁴⁸ But, once again, the radical view bumps up against intuition. Suppose that Lucy and Louie have a disposition with identical intrinsic properties. But Lucy's disposition systematically brings about good outcomes while Louie's does not. Louie has no reason to think his disposition has gone awry; perhaps an evil demon secretly brings about a very horrible outcome every time Louie acts. Intuitively, if Lucy and Louie act wrongly from their shared disposition, and Louie has no reason to suspect there is an evil demon, there is no difference in their level of blameworthiness. But the radical skeptic strangely predicts that there is some difference in blameworthiness. Lucy's act bears extrinsic moral value and Louie's does not, so it predicts that the two are not equally blameworthy.⁴⁹ The symmetry in our

⁴⁶ I assume here that our practices of praising and blaming track relevant differences in moral value.

⁴⁷ She may also try to accommodate the judgment by claiming that the *acts* are equally worthy, but we judge *agents* who act for the right reason to be more praiseworthy. Occasionally, Driver is tempted by this thought. In an attempt to make sense of similar discrepancies in our judgments, she admits that "a person's states of mind will matter when it comes to apportioning praise and blame to the person herself" (2013: 170). But it is a misrepresentation to say that our judgment merely concerns the agents and not their acts. The judgment is that the agent who acts for the right reason deserves more praise *for acting as he did* than an agent who acts for the wrong reasons does *for acting as she did*. It is not merely that the former is a better person than latter. If this were the object of our judgment, then presumably the judgment would be insensitive to whether the agent do *anything at all*. After all, one person may have been a better person than other even if he failed to act on this occasion. But, of course, we wouldn't praise the agent if he had failed to act as he did. So it seems the judgment must be directed toward the acts and not merely toward the agents in general. Driver is more committed to the response involving extrinsic value. She claims that states of mind "matter because they cause good outcomes. They cause such outcomes not infallibly, of course, but systematically or nonaccidentally" (Driver 2013: 170).

⁴⁸ For examples of noninstrumental extrinsic value, see Korsgaard (1983) and Langton (2007). For example, a paintbrush may have noninstrumental value in virtue of an extrinsic or relational property, e.g. being once owned by Picasso.

⁴⁹ Driver expects the 'systematically' caveat to protect against cases of luck: "...it is important to note that 'systematically' does a lot of work in dealing with the moral luck problem. It rules out flukes in the actual world conferring value. Thus we are still justified in ascribing virtue to agents who, through mere bad luck, were not able to produce good" (2001: 82). But, as I hope is obvious, Louie's bad fortune is *not* avoided by the requirement that the disposition cause good outcomes systematically.

judgments about Louie and Lucy makes sense if the moral value of motives is not purely extrinsic. If motives bear intrinsic moral value, two acts that bring about drastically different consequences can sometimes have equal moral value or disvalue.⁵⁰

It seems clear that radical skepticism inherits none of the intuitive plausibility that moderate skepticism enjoyed. If one intends to be a radical skeptic, there are unattractive consequences that must be explained away. And so, minimally, moderate skeptics should hesitate before retreating to the more radical stance.

7. Virtue Theory

Philosophers interested in virtue are often challenged to demonstrate the intrinsic worth of acting from virtue. In one critique, Brad Hooker writes: “[there] is no dispute among the main moral theories that the virtues are instrumentally valuable... [but] virtue ethics is often said to go beyond its rivals by claiming that virtue is not instrumentally but also *intrinsically* (i.e. non-instrumentally) valuable” (Hooker 2002: 22). He goes on to say that virtue theory has failed to motivate this core thesis.⁵¹ Crisp too notes that virtue ethicists face some burden to explain why virtue bears intrinsic value, although he suggests they may be able to meet this demand by focusing on the categorical bases of virtuous dispositions (2015: 271).

This value problem is one of the more trenchant obstacles facing the development of

⁵⁰ Of course, we needn't conclude that consequences do not matter at all. Some think that the problem of resultant luck blocks this view: a reckless driver who unluckily kills someone may be more blameworthy than similarly reckless driver who luckily manages to avoid hurting anyone. See Nagel (1979) for a seminal discussion of moral luck. Zimmerman (1987) uses the term 'resultant' luck. I am happy to grant that the moral value of motives is not purely intrinsic. It is open to anyone but the radical skeptic to claim that motives have moral value in virtue of intrinsic properties and extrinsic ones, like the property of having actually caused an act with horrible consequences

⁵¹ In other work, Hooker appears to grant that the virtues are intrinsically valuable (e.g. Hooker 2000: 36-37). But he does so in a way that treats occurrent attitudes as *primary* bearers of intrinsic moral value (2000: 34-36)

virtue ethics. There has been no successful attempt to address it.⁵² My response to the value challenge is to reject it altogether. I claim that it is a mistake to think that virtue theorists face some *special* burden to demonstrate why a virtuous disposition is intrinsically valuable. In the preceding sections, I argued that virtually everyone—excluding a radical skeptic—must agree that the virtues bear moral value intrinsically. That is, nearly everyone is committed to the intrinsic moral value of acting from virtue. Virtue ethics faces no unique challenge in this regard.

Even more, if moderate skepticism is no longer a viable position, other familiar challenges to virtue theory look similarly suspicious. For example, a standing challenge to virtue ethics rests on experimental data that seems to undercut the idea that human beings have character traits.⁵³ The data shows that people exhibit only “highly contextualized dispositions or ‘local’ traits” (Doris 2002: 64). But, when someone possesses a trait, they are disposed to perform “trait-relevant behavior across a wide variety of trait-relevant situations” (Doris 2002:

⁵² In a recent essay, Brad Cokelet tries to address the challenge on the skeptic’s terms. Unlike me, Cokelet grants that there are two ways of acting well, i.e. acting from virtue and acting from the right kind of motive. He says: “[for] the sake of argument, I am going to simply grant that an agent’s occurrent states do not logically or metaphysically depend on more general dispositional facts about her character” (2015:234-5). The moral value of acting from virtue rests in what he calls the ‘personal values conception’ of character. Consider Randy, who does not value his son or their relationship very much, but on one occasion, attends his son’s soccer game in order to show support. Cokelet insists that “the value and worth of Randy’s showing up (for good reasons) will wax and wane with his overall values; his act of showing up will ‘mean less’ and be less good and praiseworthy, if he is a bad dad than if he is a good one” (2015: 240).

One concern is that Randy’s case may turn on special features of the kind of act he performs. As Cokelet notes, the value of Randy attending his son’s game is largely symbolic, so it may not be striking that the value of his act depends on whether it expresses Randy’s deeply-held values. In cases like these, deeply-held values may be a precondition on having the right kind of motive in the first place. The account struggles when it is extended beyond acts of this sort. Cokelet claims that the value of giving \$20 to someone in need is diminished if the agent undervalues the loss of \$20. He says: “if the person has been inducted into a strange religious cult, disvalues having money in her wallet, and is therefore disposed to give her money away to anyone who asks, then her act is less generous than it would be if she values what she gives away” (2015: 242–3). But this explanation has an unsavory consequence. It seems to follow that an agent can increase the moral value of her actions by valuing morally irrelevant things. If Cokelet is right, a minor act of charity may gain moral worth simply because the agent values money a great deal. Likewise, a minor selfless act is made more morally valuable because the agent is grossly selfish. But that can’t be right. Cokelet’s attempt to answer the challenge on the skeptic’s terms seems to fail.

⁵³ See Doris (2002: 15–62) for a summary.

18). So, according to what is now known as the situationist critique, people do not have traits as they are traditionally conceived. This result is supposed to be devastating for virtue ethics because, unlike other moral theories, virtue ethics puts emphasis on the virtues and vices.

There are a number of places to put pressure on this critique. Virtue ethicists have been especially keen to point out that the use of ‘trait’ in the studies may not match what philosophers mean when they talk about virtue.⁵⁴ But, to those who take the data seriously, this reply seems like a flat-footed denial that virtue theory is “subject to charges of empirical inadequacy” (Doris 2005: 663). As a result, this debate has reached an impasse: some think that the data requires substantial revision to virtue theory and others believe it has no bearing on moral theory whatsoever.

A more judicious reply has been ignored. As I see it, the issue raised by the situationist data is fully general. With the exception of radical skeptics, everyone advances a conception of the virtues and, to the extent that people can act well, everyone is committed to the possibility of acting from that state. And so, *everyone* must measure their theory against the situationist data. There is, once again, no unique problem for virtue ethics.⁵⁵

In fact, it is strange that virtue ethics is especially targeted by situationism. As noted earlier, a simple utilitarian claims there is a single intrinsically-valuable virtue, i.e. the disposition to maximize overall utility. Unless a utilitarian is prepared to retreat to radical skepticism, situationism presents a crushing problem for any view like it. If the situationist data shows anything, it shows that the utilitarian’s kind of virtue is rare. On the other hand, virtue

⁵⁴ For this reply, see Swanton (2003: 30), Sreenivasan (2002), Kamtekar (2004), and Annas (2005: 638).

⁵⁵ To his credit, John Doris sees the general nature of the problem, acknowledging that “[character] is not the proprietary interest of Aristotelianism” (2002: 107). Still, the situationist data has been mostly directed at virtue ethics.

ethics is typically associated with some kind of value pluralism or even particularism.⁵⁶ So, for some virtue ethicists, there are many right-making reasons and thus fairly narrow virtues. Doris should think these virtue ethicists were *already* moving in the direction that the situationist data requires. In light of the data, Doris himself wants to replace attributions like ‘courageous’ with attributions like ‘sailing-in-rough-weather-with-one’s-friend-courageous’ (2002: 115). Perhaps he thinks that virtue ethicists have not gone far enough in narrowing the virtues. But, at the very least, he should agree that some versions of virtue ethics fare better than their monist competitors.

8. Conclusion

Occasionally, philosophers draw distinctions where they should not. It is a mistake to think that there are two *species* of ways to act well on any given occasion, i.e. acting from virtue and acting for the right reason. I am not saying that there is only a single way of acting well on any given occasion. In a given case, morality might permit being moved by prudence, i.e. the fact that I am in harm’s way, or benevolence, i.e. the fact that others are also in harm’s way. But these ways of acting well do not divide into two general *categories*: those which manifest virtue and those which merely show responsiveness to right reasons.

This bifurcated view of acting well rests on a mistake. The view is sometimes propped up by the observation that agents can act for the right reasons but ‘out of character.’ The version of this claim that the skeptic needs, namely that *R* states and *V* states are different dispositions, is not available to her. Even so, there are many other reasonable things that acting well but ‘out of

⁵⁶ See, for instance, Stocker (1992) and Swanton (2003)

character' could mean. It could mean, for instance, that the agent acted well even though no one expected it from her or it was not predicted by her track record. To the extent that it is psychologically possible, it may mean the person momentarily manifested virtue. Any of these interpretations are consistent with what I have said. All I have ruled out is that there are two ways of acting well on any given occasion.

By denying the coherence of moderate skepticism, I also deny the coherence of the examples typically used to motivate the view. Recall Crisp's example of Ronnie, a malicious gangster who saves a stranger on an occasion for the right reason (2015:14). Crisp claims that Ronnie acts for the right reason but does not act from a virtuous disposition. I have ruled out this possibility; either he acts for the right reason and from the relevant virtue, or else he fails to do both. Still, there are several ways of explaining what is happening in Ronnie's case. For example, if it is psychologically possible to manifest a broad disposition momentarily, and the relevant virtue is a broad disposition, then Ronnie may indeed act well. Or, alternatively, Ronnie may act well despite manifesting only a narrow disposition if the relevant virtue is actually quite fine-grained. Further progress requires a more substantive theory of the virtues and the limits of psychological change.

Whenever someone says there is only one category where others thought there were two, the claim will appear suspicious. But I think there is a good explanation of why the error cropped up in the first place. One might come to think there are two general ways of acting well by mismatching our theories of right reasons and character traits. Suppose one combines a *particularist* account of right-making reasons with a *pluralist* account of character traits. A pluralist account of character traits will say that there are *several* rather broad virtues, while a particularist account of right-making reasons will say that there are *many* quite narrow right

reasons. This yields the asymmetric relationship that the skeptic needs. But, since no moral theory can be both pluralist and particularist in this way, there is no asymmetry.

Once the bifurcated view is put aside, moderate skepticism about virtue collapses. It is no longer possible to say that the virtues are not intrinsically valuable while good motives are. As a result, the only coherent skeptical position is quite radical. But, since radical skepticism faces serious difficulties itself, there is little room for skepticism about the virtues. If the virtue ethicist wants to build an account of moral worth with virtue as the primary bearer of value, the skeptic's value challenge should not stand in her way.

CHAPTER TWO

Virtue and Excuse

Some wrongdoings are excused. When someone offers an excuse for her wrongdoing, she admits that her act was wrong, but pleads that she is not fully blameworthy for performing it.¹ Cases of excused wrongdoing highlight a divergence in our judgments about the wrongfulness of acts and the blameworthiness of the agents who perform them.

Excuses in this sense are both familiar and diverse. Some excuses involve *ignorance*: a fiancée who mistakes cyanide for sugar may be excused for poisoning her partner's coffee, as long as her mistake is itself innocent. Other excuses involve brief *incapacity*: an employee who is provoked to anger, and temporarily loses his wits, may be excused for making a hateful remark about his antagonist. Even more kinds of excuse appeal to mere *accidents*: a father who endeavors to comfort his daughter may be excused if he inadvertently reminds her of a painful memory. The list could go on.²

My focus in this chapter is an especially puzzling class of excuses. These are not grounded in ignorance, incapacity, or mere accident. In cases of *duress*, clear-eyed agents decide what to do, and act wrongly, but they nonetheless appear at least partially blameless. Wrongful action under duress is regrettably real. Consider the case of Johann Vican, a prisoner at

¹ In this regard, my use of 'excuse' follows J. L. Austin; in cases of excuse, he says, "we admit that [the act] was bad but don't accept full, or even any, responsibility" (1956: 2). Also see Kelly (2013).

² Baron (2006) points to excuses grounded in civility (5) and excuses offered for future conduct (30), and others that are sometimes called exemptions (31–31). Tannenbaum (2014) discusses a class of partial excuses she calls 'mere moral failures.' Also see Husak (2005) and Duff (2006).

Auschwitz who was recruited to participate in the murder of fellow prisoners by Nazi guards.³ Despite having no sympathy for the Nazi program or hatred for other prisoners, he carried out horrid crimes under threat of receiving punishment himself. One might quibble about the details of Vican's actual case. But it seems that an agent resembling Vican—one who knowingly and soberly performs a wrong act in terrible circumstances like his—may be at least partially excused for his wrongdoing. Yet, if this excuse is not grounded in rational incapacity, mistake, or ignorance, what explains it?

The most common response is what I call *norm-splitting*. On this approach, cases of duress highlight a significant dissonance in moral norms. In schematic form, these explanations proceed like this: Vican falls short of the moral norm governing proper conduct—that is, he acts wrongly—but, at the same time, he meets some *other* moral standard and so does not warrant full blame. Each version of the strategy offers a different proposal about the identity of this other moral standard. Some would propose that Vican meets the standard of morally permissible concern, while others would claim that he meets the moral norm governing how one ought to be disposed to act.

In the first part of this chapter, I argue that the norm-splitting strategy is misguided. Either the moral norms do not come apart in the way required or, if they do, the explanation fails to identify a sufficient condition for excuse. Even so, norm-splitting can seem inevitable. What alternative could there be?

In the second part of this chapter, I offer one: agents who act wrongly under duress are *fully but fallibly virtuous*. Since I offer this account as an alternative to norm-splitting, I cannot secure the distinction between full and infallible virtue by pointing to discordant moral norms.

³ See *United States v. Vican, Case No. 000-Flossenburg-3*. The case is discussed in Heim (2013: 181–184)

Instead, I secure this crucial distinction by offering a psychological theory. A virtuous person is disposed to act well, but (I claim) she may act wrongly on a given occasion because her psychological disposition tolerates some failures to manifest.⁴ When someone fails to manifest her virtuous disposition in a tolerable case, she is fully but fallibly virtuous. I argue that this proposal can explain excuse by duress without succumbing to the normal pitfalls of psychological and character-based theories of blameworthiness.

In section one, I describe the crucial features of wrongdoings that are excused by duress. In section two, I argue against the strategy of norm-splitting. In section three, I spell out the distinction between full and infallible virtue. In section four, I put the distinction to work in explaining why acting under duress excuses wrongdoing.

1. Duress as Excuse

Vican is an extreme example, but his case is not unique in many respects. Others have offered the case of Dražen Erdemović as an example of excuse by duress.⁵ Erdemović was a nineteen-year-old soldier during the Bosnian War in 1995. In July of that year, he participated in the massacre of over one thousand prisoners. Erdemović initially refused to take part, but ultimately relented under threat that he would be killed if he did not participate. He had no interest in the conflict and had been previously dismissed from service for helping prisoners escape abuse. Following the massacre, he deserted the service as soon as he could and helped bring the

⁴ Crucially, my claim here is not that morally permissible dispositions sometimes issue in bad conduct; this claim would amount to a norm-splitting explanation. My proposal is that virtuous dispositions, which never *issue in* bad conduct, may nonetheless subsist *in spite of* some kinds of impermissible conduct. As I will show, this subtle difference is crucial.

⁵ Rosen (2014: 71–73) offers a more complete description of the case. Also see Turns (1998) and Brooks (2003).

atrocities of the war to international attention. When we bear in mind the terrible circumstances in which he acted, Erdemović too seems to deserve at least partial excuse for his wrongdoing. Excuses like these display a few crucial features:

First, in these cases, the action performed is morally wrong.⁶ Nearly everyone should agree that duress may sometimes *justify* behaviour that would otherwise be impermissible. Consider a bank teller who turns over his client's money under threat by an armed robber. Although the teller acts under duress, he doesn't act *wrongly* by handing over the money. His behaviour is justified by the fact of duress; when you are threatened by physical harm, it is permissible to give up valuables entrusted to you in order to protect your life. As I see it, things are different for Erdemović and Vican. Unlike the teller, Erdemović and Vican act impermissibly in their respective circumstances. Given the gravity of their acts, and the sheer amount of death and pain they produce, it seems Erdemović and Vican should not have acted as they did.⁷

Of course, there is room to debate whether Vican and Erdemović really do act impermissibly. But this shouldn't concern us here. There must be some modification of their cases that makes their behaviour obviously impermissible. Claire Finkelstein suggests modifying cases like these by making the consequences of the acts even worse (2002: 331). At some point in this intensifying process, it will become clear that Vican and Erdemović should give up their lives instead of acting as they are directed. Still, Finkelstein thinks, these agents will be at least partially excused if they act under genuine threat of punishment or death.

⁶ This is to say that Erdemović and Vican ought not to act as they do. The sense of 'ought' is both objective and subjective, since I assume that they have ample evidence about what their actions will bring about. I assume they violate an *all things considered* obligation, not a *pro tanto* one.

⁷ With a modified version of Erdemović's case in mind, Rosen concludes: "*if there is ever a case in which a person might be morally required to die rather than kill the innocent, this is it, or this is close to it*" (2014: 78 his emphasis).

Second, in cases of excuse by duress, the agent acts deliberately. That is, the agents I have in mind *decide* to act as they do. Of course, it may have been the case that Vican and Erdemović were so paralyzed by fear that they briefly lost their capacity for reflective self-control.⁸ But this needn't have been the case for them to deserve some kind of excuse. An agent who is excused by duress may have deliberated about what to do, at least as much as any of us do in everyday life. The ground for their excuse does not lie in the fact that they did not genuinely choose to act.⁹

Notice that, if agents like Vican and Erdemović deliberately perform a wrong act, they are not *compelled* in one familiar sense of the term. Consider Gary Watson's influential discussion of compulsion and weakness of will.¹⁰ For Watson, weak-willed people "give in to desires which the possession of the normal degree of self-control would enable them to resist" but compulsive people give into desires "such that the normal capacities of resistance are or would be insufficient to enable the agent to resist" (1977: 48–49). He understands 'normal capacities of self-control' as those capacities which are acquired in ordinary socialization and possessed by most adults (1977: 50–51). It is tempting to assimilate our cases of duress to cases of compulsion in Watson's sense. However, if duress is expected to provide an authentic excuse for Vican and Erdemović, this temptation should be resisted. To see why, consider the following:

As Watson points out, weak-willed and compulsive agents are both constitutionally

⁸ I borrow 'reflective self-control' from R. Jay Wallace. For Wallace, the powers of reflective self-control involve the ability to discern and apply moral reasons, as well as regulate one's behavior in the light of those reasons. See Wallace (2006: 86, 157)

⁹ I grant that the Vican and Erdemović's options have been impermissibly limited. In their respective circumstances, they are not afforded to option of *resisting wrongdoing and avoiding execution*. But the fact that their options are limited in this way does not explain their excuse because, in general, agents with impermissibly limited options can be blameworthy for what they do (cf. Frankfurt 1969). A soldier who has no qualms about murdering civilians may have similarly limited options—say, he knows that his commander will kill him if he does not comply—but he still appears morally blameworthy because the threat plays no motivational role. So the explanation of excuse by duress must point to something more than impermissibly limited options.

¹⁰ Watson (1977)

incapable of doing otherwise at the time of acting.¹¹ These agents differ because their incapacity is explained in different ways: the weak-willed person is incapable *because* she lacks normal self-control, but the compulsive person is incapable *in spite of* her normal capacity for self-control. This kind of incapacity suggests that, strictly-speaking, the compulsive person does not behave wrongly in acting as her compulsion requires. According to Watson, the compulsive agent cannot perform any other behaviour when her compulsion is activated. Granting the ‘ought implies can’ principle, it follows that the compulsive person cannot be obligated to perform an alternative behaviour. As a result, the compulsive person does not violate a moral obligation by behaving as her compulsion requires.¹² This is an important departure from our cases of interest. In cases of excuse by duress, the agent is *morally required* to behave differently; at least on modified versions of their cases, Vican and Erdemović ought to refrain from acting as they do. Since the compulsive person is not a genuine candidate for excuse, our cases of duress cannot be assimilated to cases of compulsion in Watson’s sense.

Third, in cases of excuse by duress, we blame agents less because they are less worthy of blame. This is to say that we do not merely *pardon* them. The difference between excusing and pardoning can be hard to see in any given case, but the two acts can be conceptually separated.¹³ In order to pardon someone, one must judge that the person is blameworthy and yet choose to withhold blame. If I am late for a meeting for totally foreseeable reasons, I

¹¹ He says that a weak-willed person must be constitutionally able to meet standards of normal self-control, but may be “unable at the time of action to resist” (Watson 1977: 54). Similarly, a compelled person “is motivated by a desire (or ‘impulse’ or ‘inclination’) that he or she is unable to resist” (1977: 43). See Hieronymi (2007: 110).

¹² Of course, by this reasoning, the weak-willed person also does not violate a moral obligation when she behaves badly. However, as Watson notes, the weak-willed person is blameworthy nonetheless (1977: 53). Although she does not violate a requirement in acting as she does, she does violate a requirement in failing to cultivate a normal capacity for self-control.

¹³ I use ‘pardon’ as Jeffrie Murphy uses ‘forgiveness.’ He writes: “We may forgive only what it is initially proper to resent; and, if a person has done nothing wrong or was not responsible for what he did, there is nothing to resent” (Murphy & Hampton 1988: 20). Also see Zaibert (2012)

am fully blameworthy for my lateness, but my kindhearted colleagues may not blame me because they decide to let my tardiness pass. I take it that agents like Vican and Erdemović are excused rather than pardoned in this sense. We do not merely *choose* to free them from blame, but rather we judge that Vican and Erdemović *deserve* reduced blame.

Lastly, I assume that the blamelessness of Vican and Erdemović is not explained by our lack of *standing* to blame. Many conditions can apparently undercut our standing to blame another person for a wrongdoing, including the fact that we would have done the same thing if we were in the person's circumstances.¹⁴ But even someone with standing to blame should excuse those who act wrongly under duress. Consider Erdemović's comrade, who resists taking part in the slaughter of prisoners, but miraculously survives subsequent execution attempts by his superiors. Since it is clear that this comrade would have given up his life to avoid wrongdoing, he certainly has standing to blame Erdemović. Yet, it seems to me that even this comrade shouldn't fully blame Erdemović. Our task now is to explain why this is so.

2. Norm-splitting

One's immediate response to duress is often quite simple: we excuse Vican and Erdemović because it is too difficult to resist acting as they did in their respective circumstances. Why can't this explanation suffice?

Taken on its own, the simple explanation is no explanation at all. Resistance is clearly not too demanding from the guise of the moral norm regulating how we ought to act. As we said, Vican and Erdemović act impermissibly. At this point, norm-splitting begins to look attractive.

¹⁴ See Watson (1987: 239–254), and Wallace (2011: 326).

A norm-splitting explanation aims to identify a moral standard or norm according to which resistance *is* too demanding. If one could identify this other moral norm, it might provide an explanation of why agents who act under duress are excused. But what could it be?

In a case study of Erdemović's case, Gideon Rosen attempts to fill out the strategy described above. He directs our attention to the norm concerning what it is morally permissible to care about (2014: 86). Rosen suggests that agents like Vican and Erdemović fall short of the moral standard of permissible conduct by acting as they do, but they meet the standard of permissible moral concern. Morality tells us, so to speak, that we are required to act in a particular way but it is permissible to be unconcerned about acting in that way.

Rosen admits that the proposal is merely a conjecture.¹⁵ But I suspect it will appeal to developed accounts of blameworthiness as well. Nomy Arpaly and Timothy Schroeder, for instance, understand some blameworthy acts as manifestations of moral indifference, which they claim is a matter of having an insufficiently "strong intrinsic desire for the right or good, correctly conceptualized" (2014: 162). They do not discuss excuses like duress, but one can see how a norm-splitting explanation might find its place in their theory. Presumably, the agent who acts wrongly under duress would have acted permissibly if he had a stronger desire for the right or good. So his actual desire for the right or good is not strong *enough* in the sense that it fails to motivate him to act permissibly. But, if he is afforded an excuse, there must be some other sense in which his desire *is* strong enough, i.e. some other moral standard that he does meet. Morality may again speak with two voices: it is impermissible for Erdemović to kill the civilians, but it is permissible for him to be so indifferent to morality that he participates in the massacre.

When confronted with cases of duress, I suspect that Arpaly and Schroeder will need to

¹⁵ Rosen (2014: 85–86)

appeal to an explanation like the one above.¹⁶ As they say, the “more things one puts ahead of the right or the good the greater one’s moral indifference” (2014: 169). Agents like Erdemović and Vican put their lives ahead of the right or good, so they show at least some degree of moral indifference. To provide these agents with a genuine excuse, Arpaly and Schroeder seem required to say that Erdemović and Vican have a morally permissible level of indifference. This explanation follows Rosen’s norm-splitting suggestion.

This kind of norm-splitting sits awkwardly with moral practice.¹⁷ It would be strange if morality *required* an action but *permitted* disinterest in performing it. One reason that this claim seems so bizarre is that there is a supposedly reliable connection between moral concern and moral conduct. One expects that people who genuinely care about the things morality says one should care about—e.g. the suffering of others—will act (or be disposed to act) as morally requires. If the norms governing conduct and concern do not share a common ground, this expectation should be somewhat baffling. But, intuitively, there is no puzzle here: the norms governing concern and conduct *do* share a common ground, so it is perfectly ordinary to expect that moral concern will correlate with moral conduct. If we accept that the two norms come apart, this straightforward explanation becomes unavailable.

One might reply by pointing to *supererogation*. Although supererogatory acts are morally

¹⁶ In other cases of excuse, Arpaly and Schroeder appeal to the fact that non-rationalizing features will play a big role in explaining what the agent does. In discussing the famous Milgram experiments, they say that a closer look “suggests practical irrationality rather than moral indifference” (2014: 168). But this analysis doesn’t extend to the cases at hand. In cases of excuse by duress, agents are rational and clear-eyed, and there is a genuine decision made.

¹⁷ Marcia Baron (2014) raises a separate worry about this version of norm-splitting. She asks whether insufficient concern is a necessary condition on blameworthiness. If not, the fact that Vican and Erdemović have a sufficient level of concern does not explain their excuse. Any explanation of their excuse should point out a *necessary* condition on blameworthiness that they fail to meet. Baron’s main counterexample is a case of *incompetence* where the agent does not manifest insufficient concern but still appears blameworthy (2014: 93). A patient dies during surgery because the tube providing her oxygen disconnects from the ventilator. Her anesthesiologist did not notice until an alarm signaled some kind of trouble. The doctor focused his attention on helping the patient, but because he hadn’t noticed that the problem was a disconnected tube, he made all the wrong decisions. Baron submits that the anesthesiologist is blameworthy for his behavior, at least on some ways of spelling things out.

good, it is permissible to be unmotivated to perform them. Typical examples include acts of heroism, e.g. the act of saving a stranger from a burning building at great personal risk. Intuitively, it is permissible to not care enough about the lives of strangers to run into a burning building to save them. But, at the same time, the heroic act is obviously morally good. Doesn't this show that the norms governing action and those governing concern come apart?

There are reasonable doubts about whether supererogation is possible.¹⁸ But, putting those doubts aside, the possibility of supererogation merely tells us that there are morally *heroic* acts that one may be unmotivated to perform. In order to explain excuse by duress, a more radical claim is needed. One must claim that there are morally *required* acts that one needn't be motivated to perform. The appeal to supererogation does not establish this more radical discrepancy in moral norms. To the extent that supererogation makes the radical schism more palatable, it also makes it reasonable to abandon the idea that Vican and Erdemović act impermissibly. This means that an appeal to supererogation will amount to giving up on a crucial feature of our cases.

It is possible that Rosen proposed the wrong norm in his version of the strategy. Hume's remarks about blameworthiness inspire a different version of norm-splitting.¹⁹ Hume claimed that wrong acts are not blameworthy unless they are rooted in morally poor dispositions. Several philosophers have endorsed this principle, occasionally touting its ability to deal with excuses like duress.²⁰ On this version of norm-splitting, Vican and Erdemović act impermissibly, but they are excused because their act is not rooted in a morally poor disposition.

It may not be obvious that this explanation is a version of norm-splitting, so I should spell

¹⁸ See, for example, Raz (1975).

¹⁹ See his *A Treatise of Human Nature*, book II, part III, sec. II.

²⁰ See, for example Brandt (1958: 3–29), Hobart (1966), Bayles (1982), Vuoso (1987), and Finkelstein (2002). For discussion, see Sher (2002) and Sher (2006: 18–32)

it out in more detail. One might have thought that the only morally acceptable dispositions are dispositions that exclusively issue in permissible conduct. The proposal above begins by denying this apparent truism. It proposes that there are some morally acceptable dispositions that issue in impermissible conduct in rare or unlikely cases.²¹ Further, the proposal continues, when someone acts impermissibly from one of these so-called acceptable dispositions, they are excused for their wrongdoing. Vican and Erdemović are cases of this sort. They act impermissibly, but they do not manifest an unacceptable disposition by the lights of the proposed explanation.

This version of norm-splitting faces a familiar counterexample.²² Some people who act wrongly from an acceptable disposition (by the proposal's lights) will *not* deserve an excuse. George Sher describes an agent named Alphonse, who is "a generous and empathetic person" and who "tries scrupulously to do the right thing" (2006: 26). Alphonse has a moral blind spot: he judges himself extremely harshly for his rare lapses in moral perfection, and when judging himself, he tends to behave "in a spirit of moral self-defilement" (2006: 24). Sher goes on to describe a particular episode in Alphonse's life: on one occasion, Alphonse believes that he has made a moral mistake and begins to morally self-destruct by making "a cruel remark of precisely the kind that he most abhors" (2006: 23–24).

Although Alphonse is disposed to act well in nearly every circumstance, he is disposed to act poorly on the exceptionally rare occasion that he makes a moral mistake.²³ By the lights of the proposed explanation, then, Alphonse acts wrongly from a morally acceptable disposition.

²¹ I use 'morally acceptable' because it is the most natural way to describe these imperfect yet morally permissible dispositions. The thesis can be rephrased using 'good' or 'permissible.' For example, 'there are morally good dispositions that issue in impermissible conduct' or 'there are morally permissible dispositions that issue in impermissible conduct.'

²² George Sher (2006: 29) discusses this objection to character theories of excuse in general. Also see Michael S. Moore (1997: 579) and Smith (2008b).

²³ Sher notes that "the dispositions that come together to produce Alphonse's cruel remark... are not themselves dispositions to act cruelly; and neither (we may suppose) would their interaction lead Alphonse to act cruelly under many other circumstances" (2006: 24–25).

But, if acting wrongly from an acceptable disposition is sufficient for an excuse, it seems Alphonse deserves an excuse for making the cruel remark. This is the wrong result; Alphonse is blameworthy for making his cruel remark. People will often be blameworthy for acting badly on the basis of their moral blind spots, regardless of how rarely they are actually granted the opportunity to express them. This version of norm-splitting fails because there are acceptably disposed people by its own lights who are blameworthy for what they do.²⁴

I don't see any other way to fill out the norm-splitting strategy.²⁵ At one point, Rosen suggests a different kind of explanation. He proposes that it is inappropriate to have attitudes of resentment toward Erdemović because, as he puts it, "Scanlon-style social distancing would not be warranted given the attitudes manifest in his act" (2014: 89). To socially distance oneself from another a human being in Scanlon's sense, one must withdraw some of the trust and willingness to cooperate that we ordinarily extend to others, even strangers.²⁶ Blaming Erdemović is inappropriate because "he [is] entitled to a seat at the table, that is, to ordinary sociability" (Rosen 2014: 88).

The proposal is peculiar because distancing oneself from another person *just is* part of blaming that person in Scanlon's view. For him, "to blame a person is to judge [that the action shows something about the agent's attitudes toward others that impairs the relations that others

²⁴ One can of course deny that Alphonse is acceptably disposed in the relevant sense. But this is inconsequential. In response, Sher can make Alphonse's moral blind spot *exceedingly* rare; perhaps he is disposed to act poorly in exactly one case. If there are some acceptable dispositions that issue in impermissible conduct, surely this extreme case is one of them. Yet, the same point will apply: if this exceedingly rare circumstance comes to pass, Alphonse will be blameworthy for what he does and this norm-splitting explanation will declare that he is excused.

²⁵ Sher's replacement for Hume's view doesn't appear to help explain excuses like duress. For him, an act is not blameworthy unless the act is "traceable to the interaction of a complex subset of the desire, beliefs and fine-grained dispositions that together make the possessor the person he is" (Sher 2006: 49). As far as I can see, Sher's condition on blameworthiness may be met in cases of duress, since the desires that motivate Vican and Erdemović to protect their lives may indeed be core to their personal identities. At the very least, it seems one needs to say why the desires that motivate them are *not* part of their personal identity. After all, we have assumed that Vican and Erdemović act on these self-protecting desires with all their rational capacities intact.

²⁶ Scanlon (2008: 140–141). See also Sher (2013).

can have with him or her] and to take your relationship with him or her to be modified in a way that this judgment of impaired relations holds to be appropriate” (2008: 128- 29). So, if one adopts Scanlon’s view of blame, the proposal is circular. As a result, the proposed explanation cannot be neutral between analyses of blame. One might have hoped to explain excuse without taking a stand on what constitutes blame. In any case, as Rosen might agree, this explanation is not really a satisfactory stopping point. Resentment may be inappropriate because social distancing is inappropriate, but why is distancing inappropriate in response to wrongdoing performed under duress? Why is it inappropriate to revise one’s relationship with Erdemović and Vican in response to what they did? It seems the proposal merely *reframes* the hard question about duress without answering it.

None of the proposals discussed so far have provided a satisfactory explanation of excuse by duress. Their failure, I suspect, has to do with their focus on identifying some discrepancy in our moral norms. In the remainder of the chapter, I take a fresh look at how one might explain excuse by duress without norm-splitting.

3. Fallible Virtue

According to a familiar picture, virtue comes in degrees. Agents who fall on one end of the spectrum are completely virtuous and agents who fall on the other are completely unvirtuous (or vicious, if you prefer).²⁷ Barring ignorance, those who are completely virtuous always do what morality asks and those who are completely unvirtuous never comply. I propose a shift away from this picture of virtue. In particular, I propose an account that recognizes the distinction

²⁷ See, for example, Annas (2004)

between *full* and *infallible* virtue. Let me begin by spelling out the distinction I have in mind.

Full virtue requires that the agent possess the psychological disposition to act well. More precisely, it is a matter of having the beliefs, desires, hopes, aversions, tastes, and knowledge that ground a disposition to act well. There is room for a view that is psychologically rich, ascribing many complex cognitive and emotional states to the virtuous person, and there is room for a view that is psychologically meager, ascribing only want and aversion to her.²⁸ Virtuousness in this sense may come in degrees; a fully virtuous person is disposed in exactly the way moral theory prescribes, but a less fully virtuous person matches this disposition incompletely.

Infallible virtue requires that the agent always acts permissibly (unless she is mistaken about some matter of fact). If full and infallible virtue come apart, then it is possible to always act permissibly, and so be infallibly virtuous, but lack a disposition to act well, and so fail to be fully virtuous in my sense. I take it that this possibility is real but quite unlikely.²⁹ Given the psychological connection between one's dispositions and one's behavior, it is quite unlikely that someone who always does the right thing will lack the relevant kind of disposition. Yet it is conceptually possible. Imagine a hapless but evil person who always tries to do the wrong thing, but unwittingly always ends up doing the right thing. This person may always act permissibly, but (we may assume) he lacks any of the beliefs, desires, and knowledge that usually

²⁸ Julia Annas (2011) provides a psychologically rich picture of virtue. Richard Brandt (1970) gives a notably less complicated account.

²⁹ This claim may be questioned by those who think permissibility sometimes depends on the agent's motive. If one believes that some acts are permissible *because* they are performed from certain motives, then anyone who is infallibly virtuous must be fully virtuous. That is, only the fully virtuous can be infallibly virtuous. See Slote (2001) and Sverdlik (2011). I am not concerned if my view of blameworthiness sits uncomfortably with this view, since the view faces difficulties when it comes to praiseworthiness and blameworthiness in general. For example, the view is incompatible with an altogether plausible principle about praiseworthiness, namely that a praiseworthy agent's motivating reasons are the right-making reasons of acting (Arpaly 2003: 67–116; Markovits 2010). If motives can be right-makers, then the praiseworthy agent's motivating reason is that she is so motivated. But this is unintelligible. In general, someone who ties permissibility and motives tightly will need to provide a total theory of praiseworthiness and blameworthiness.

characterize virtue. The hapless but evil person is infallibly but not fully virtuous.

The distinction between full and infallible virtue also implies that it is possible to maintain a virtuous disposition, but fail to act permissibly. This possibility appears to face a more serious conceptual problem. In the remainder of this section, my goal is to show that this problem can be resolved.

When someone is fully but fallibly virtuous, she is disposed to act permissibly but fails to do so. Our ordinary practice of ascribing psychological dispositions allows for this possibility. In general, dispositions survive some failures to manifest. A professional basketball player may be disposed to make a free throw when she tries. But, if the fans flash a bright light just as she tries to shoot, she may miss in spite of her disposition to make the shot. In these cases, one might say that the disposition *tolerates* the failure.

The idea that psychological dispositions may tolerate failures to manifest has been discussed elsewhere.³⁰ Notably, John Heil and C.B. Martin appeal to fallibility in order to circumvent one of Kripke's objections to the dispositional analysis of rule-following.³¹ Kripke claimed that grasping a rule cannot be a matter of having a disposition because of the possibility of mistake. According to the objection, the dispositional account cannot distinguish someone who makes a mistake in applying a rule and someone who grasps a different rule altogether. In response, Heil and Martin claim that an agent's "having acquired, and continuing to possess, the disposition does not mean that [he] will manifest it on every occasion in which its manifestation is called for" (1998: 291). In cases of mere mistake, the agent's rule-following disposition is blocked or inhibited but nonetheless still present.

³⁰ There is evidence that Thomas Aquinas endorsed the distinct possibility of fallible dispositions to act. On this point, see Kent (2012: 151–154).

³¹ Heil & Martin (1998: 291) and Kripke (1982). See also Bird & Handfield (2008)

Likewise, D.M. Armstrong appeals to fallibility to defend his claim that having an experience that *p* requires being disposed to believe that *p*. In cases of illusion, the world appears to an agent in a peculiar way, but she may withhold believing that things are how they appear. In short, Armstrong's reply is that the appearance still "pushes towards being a belief, but is held back by other contradictory beliefs" (1961: 106).

The main objection facing the idea of full but fallible virtue proceeds as follows. If full but fallible virtue is possible, then someone who is disposed to act permissibly may act impermissibly on occasion. On one hand, if a given failure is explained by a feature external to the agent—say, she is physically prevented of fulfilling her intentions—then the agent does not truly act impermissibly. After all, if she is prevented from performing an act, she cannot be obligated to perform it. But, on the other hand, if the disposition is interrupted by some feature internal to her psychology—say, her beliefs or desires—then she must lack the psychological disposition in the first place.³² So, it seems, one cannot have a disposition to act permissibly and act impermissibly.

This objection rests on a false assumption. It is plainly false that an agent must lack a given disposition if she fails to manifest it on account of features internal to her psychology. Consider a case where an agent's psychological disposition is resisted by the agent's mental behaviour. Emerson is an extraordinarily anxious person. At any mention of a performance review at his job, he is disposed to engage in a series of pathological behaviours, e.g. recalling his many mistakes, wondering intensely about the review, and deliberating about what to do if he is fired. Knowing that he is disposed in this way, he decides to rehearse a favorite song to distract himself whenever the review is mentioned. Emerson remains anxious and disposed to

³² Bird & Handfield (2008) raise a version of this objection Heil & Martin's proposal.

perform unhealthy mental acts despite his failure to do so when he rehearses the song. His anxious disposition makes his mental rehearsing intelligible; he sings to himself because he is prone to anxious spirals. It seems, intuitively, that Emerson has interrupted his disposition by performing a mental act. His failure to act is explained entirely by a feature internal to his psychology.³³

Philosophers sometimes deny that features internal to an agent's psychology can interrupt her dispositions because they are concerned about the coherence of *intrinsic finking* or *masking*.³⁴ Consider David Lewis' well-known example of finking: a sorcerer protects a fragile glass by carefully watching over it and quickly casting a spell that makes the glass unbreakable if it is struck. In this case, the glass' disposition to break when struck is finked by the sorcerer. Mark Johnston provides a familiar example of masking: a fragile glass wrapped in bubble wrap may not break when struck. In this case, the glass' disposition to break when struck is masked by the bubble wrap. Some philosophers believe it is crucial that the finking or masking property—being under the sorcerer's watchful eye or being covered in bubble wrap—is an *extrinsic* property of the glass in both cases. Suppose the sorcerer decided to “enchant the glass in such a

³³ Note that Emerson's anxious disposition does not meet the conditional or nomic duplicate tests proposed in Choi (2005: 499). The conditional test asks whether the object would manifest the disposition if the stimulus conditions were to obtain. The nomic duplicate test asks whether it is 'clear enough' that a perfect duplicate of an object (subject to the same laws) has a disposition. Emerson fails the conditional test; he would not have an anxious spiral when his review is mentioned. Presumably, a perfect duplicate of Emerson—call him Emerson*—would also resist his anxious disposition, since Emerson* shares the beliefs and desires that motivate Emerson to sing to himself. I suspect then that it is not clear enough that Emerson* has the anxious disposition either, so Emerson fails the nomic duplicate test.

But this needn't worry us. Choi's tests is limited, as he later recognized (Choi 2011: 300). Randolph Clark (2008: 515) imagines a man, John, who shares all the intrinsic properties of a strong person, but also has a strange intrinsic property, P, that would cause him to lose the intrinsic properties of a strong person if he were to touch a heavy thing. John wouldn't lift a heavy thing if he tried, so he fails the conditional test for strength. John's nomic duplicates are not clearly strong either. So, John is *not* strong by Choi's tests. But, imagine a duplicate of John—John*--who would have P removed by a wizard if he tries to lift something heavy. Not only would John* lift heavy things if he tried, but all nomic duplicates of John* appear strong too. So, John* *is* strong by Choi's test. But the only difference between John and John* is the presence of the wizard and this shouldn't bear on whether these men are strong. Yet Choi's tests deliver different verdicts about them. Also see Handfield (2008).

³⁴ See dispositional finks in Lewis (1997) and Martin (1994). Also see dispositional masks and antidotes in Johnston (1992) and Bird (1998), respectively.

way that it has an intrinsic property which would cause it, when struck, to lose the molecular structure which, under normal circumstances, is the casual basis of fragility” (Bird & Handfield 2008: 8). One might insist that a glass with this peculiar *intrinsic* property is simply not a fragile glass. In short, there is reason to think that intrinsic finks or masks are impossible.

The coherence of intrinsic finks and masks is a divisive topic in metaphysics.³⁵ But the possibility of full but fallible virtue needn’t depend on the plausibility of intrinsic masking or finking. Not every failure to manifest a disposition calls out for a finking or masking explanation.³⁶ Whether a particular failure to manifest calls out for one of these explanations will depend on the level of generality at which the disposition is specified. Consider an example inspired by Sungho Choi: people get sick when they ingest seawater.³⁷ If someone ingests just a small amount of seawater, she may not get sick. This failure is a counterinstance vis-à-vis the disposition to get sick when one ingests seawater, but it certainly does not falsify the dispositional claim; it remains true that people get sick when they ingest seawater. Furthermore, this counterinstance does not call out for a masking or finking explanation. For any dispositional claim at a specified level of generality, there are tolerable counterinstances of this sort. The possibility of this kind of tolerable counterinstance is all that is needed to vindicate the notion of full but fallible virtue.

³⁵ For skepticism about intrinsic finks and masks, see Choi (2003), Handfield (2008: 302), Handfield & Bird (2008), and Choi (2009). For defenses of intrinsic finks and masks, see Clarke (2008), Fara (2008), and Ashwell (2010, 2015)

³⁶ Michael Fara sometimes seems to take every counterinstance to a disposition to be a case of dispositional masking or finking. He says: “A piece of wood in a vacuum chamber is not less disposed to burn when heated than is its aerated counterpart...but wood won’t burn if heated in a vacuum. The masking of dispositions is such a humdrum occurrence that any adequate account of dispositional ascriptions must accommodate it” (Fara 2005: 50). But this case is not really a masking case in the technical sense. His case is one where the stimulus conditions for the disposition do not obtain. Suppose the wood does not burn because it is protected by a layer of fireproof paint. In this case, the stimulus conditions may well obtain, but the disposition does not manifest because it is masked in the technical sense.

³⁷ See Choi (2012: 1164)

Notice that tolerable counterinstances persist at every level of generality. Consider a more specific dispositional claim: people get sick when they drink at least a cup of seawater. If someone ingests a cup of seawater, but consumes it over two days, she too may not get sick. This is a counterinstance vis-à-vis the disposition to get sick when one ingests a cup of seawater. But again, it does not falsify the dispositional claim, nor does it call out for a finking or masking explanation. Consider an even more specific claim: people get sick when they drink a cup of seawater in one hour. If someone ingests a cup of seawater in one hour, but the water has an abnormally low salt concentration, she may not get sick. But again, this counterinstance needn't signal that the dispositional claim is false, or that the disposition has been finked or masked. This kind of tolerable counterinstance persists at every level of generality.³⁸

The dispositions I have called full virtues will tolerate some counterinstances. Suppose the virtue of gratitude is the disposition to acknowledge people when they have supported one's pursuits.³⁹ Jane believes she has been professionally successful because of her supporting team, so she regularly thanks and credits her team in appropriate ways. One day, Jane is tasked with giving a rare presentation to her company's board of directors. During her talk, she begins to feel as though her audience is growing bored. In an effort to speed things up, she decides to skip the prepared acknowledgement portion of her talk, though she is immediately distressed by her decision. Jane's behaviour is a counterinstance vis-à-vis the disposition identified as gratitude above. But, despite this failure, it seems to me that the relevant dispositional claim remains true; it remains true that Jane is disposed to acknowledge the people who support her.⁴⁰ As I see it,

³⁸ This kind of counterinstance is sometimes raised as a problem for conditional analyses of dispositions. Lewis (1997: 145) argues that context will set the precise stimulus conditions for any dispositional ascription, so the conditional analysis can in principle accommodate these counterinstances. See Manley and Wasserman (2008: esp. 63–66) and Choi (2011: 1169).

³⁹ W.D. Ross (1930: 21–22) seems to characterize gratitude in this way.

⁴⁰ Notice that I have not characterized gratitude as the disposition to acknowledge a person when the person has been supportive *and* one is stressed for time. Perhaps Jane's failure to act entails that she lacks this disposition (or

the judgment that Jane's disposition tolerates her failure is a verdict rendered by folk psychology. It is the responsibility of moral theory to identify the virtues at the appropriate level of generality; it is the job of folk psychology to determine how one must behave in order to count as possessing or maintaining those virtues.

I propose that the traditional picture of virtue runs things backward. On the traditional picture, one begins with a conception of infallible virtue and defines full virtue accordingly. To be fully virtuous, a traditionalist claims, one must possess the psychological disposition that does not tolerate any failures to act well. A fully virtuous person has *whatever* psychological states are necessary to ensure that one never flouts a moral requirement. My proposal is to run things the other way: begin with a conception of full virtue and define infallible virtue accordingly.⁴¹ To be infallibly virtuous, one must always manifest full virtue, i.e. one must always act well. I have no doubt that this conceptual reversal will be controversial. But, granting for the moment that the account can be developed, it promises to provide an explanation of the excuses that have been so perplexing. My next task is to show exactly how this psychological explanation might proceed.

4. Explaining Excuse

Psychological factors have been introduced to explain duress elsewhere. According to what we

some kind of intrinsic masking has taken place). But this more specific disposition that she lacks (or is masked) is *not* the virtue. The virtue is the disposition to acknowledge a person when that person has been supportive. This latter disposition may tolerate the failure, even though the former one does not.

⁴¹ This may somewhat overstate the point. Of course, one wants give a description of the virtues such that they never issue in impermissible conduct. So, in giving a description of the virtues, one needs to be conscious of how the specified disposition will manifest in conduct. On this much, I agree. But it is one thing to say that 'a conception of impermissible conduct will *inform* one's description of the virtues' and quite another to say that 'the virtues are *whatever* psychological states are necessary to require that one never flouts a moral requirement.' The latter proposal is the traditional view I aim to replace.

might call *pressure views*, wrongdoing under duress is excused because the magnitude of the threat makes it psychologically difficult to resist wrongdoing.⁴² There are several problems with the view, but perhaps the most severe is that it is too permissive. It entails that one should excuse those who feel psychological pressure when the threat is not genuine or not “important on some scale of values independent of the agent’s own” (Richards 1987: 23). The tendency is to move directly from psychological pressure views to norm-splitting. With the observation above in mind, Jeffrie Murphy explicitly concludes that, “in most cases properly called duress...what we are characterizing is...something mainly moral rather than mainly psychological” (1992: 133).

Murphy’s move may be premature. In one way, pressure views are on the right track: the experience of inner struggle is crucial when we excuse agents by appeal to duress. If, for example, Erdemović did not experience inner struggle in acting as he did, I gather our judgment that he is blameless would subside. But, the pressure view goes awry by grounding his excuse in the experience of struggle directly. The fact that explains his excuse is not the inner struggle itself, but the fact that inner struggle *reveals*, namely that he is disposed to act well in the relevant case. This is how I intend to explain the excuse afforded to agents who act under duress. Let me spell it out:

Suppose that Vican and Erdemović are fully virtuous in my sense. Of course, I do not take this starting assumption to be obviously true, but I make it in order to put the account in view; I will relax the claim in a moment. If they are fully virtuous, they are disposed in precisely the way moral theory prescribes and they have the beliefs, desires, and other mental states that ground the virtuous disposition. Vican and Erdemović act impermissibly in their respective circumstances, but our folk psychological judgment is that their failure does not extinguish or

⁴² See Richards (1979) and Murphy (1992: 131–134) for discussion.

diminish their virtuous disposition. To use the notion described before, they are fully but fallibly virtuous. In general, agents who are fully virtuous but act impermissibly are not blameworthy for what they do. So Vican and Erdemović are excused.

Is this explanation an alternative to norm-splitting? The explanation above may appear similar to the norm-splitting explanation inspired by Hume. However, there is an absolutely crucial difference. The earlier norm-splitting explanation claims that morality sets one standard for acceptable dispositions and a different one for permissible conduct; in other words, it claims that morally acceptable dispositions may sometimes issue in wrong behaviour. This causes trouble for the proposal. On my account, morality speaks univocally: the only morally acceptable dispositions are dispositions that issue exclusively in permissible conduct. An acceptably disposed person may act wrongly on occasion, but (I claim) this is *not* explained by the existence of morally acceptable dispositions that issue in wrong behaviour. Rather, it is explained by the fact that morally acceptable dispositions psychologically tolerate some failures to manifest.

One way to see the difference is to consider how my proposal deals with the counterexample to the earlier view. The norm-splitting proposal was committed to excusing people when they express moral blind spots. Someone who is disposed to act well in most cases and poorly in rare cases will be excused whenever those rare cases come to pass. This implication is fatal for the earlier version of norm-splitting and it has cast doubt on any other character-based theory of excuse. On my view, however, the only people who deserve excuse are fully virtuous. Since someone who has a moral blind spot—that is, someone who is disposed to act poorly in some cases—cannot be fully virtuous, no one who manifests a moral blind spot will be excused. The counterexample to the earlier view does not get a grip on the present proposal

This means that the view I propose *can* provide a necessary condition on

blameworthiness: an agent is not blameworthy for a wrongdoing unless she fails to manifest full virtue in my sense. My explanation of excuse by duress depends on this claim. It provides the necessary condition on blameworthiness that Vican and Erdemović fail to meet. Unsurprisingly, this condition would be trivial if not for the distinction between full and infallible virtue. On the traditional picture, full virtue is perfect compliance with the demands of morality, so it follows that *every* wrongdoing must be performed by someone who is not fully virtuous. But, since I propose that full virtue and infallible virtue come apart, there is room for wrongdoings that are performed by fully virtuous people. These wrongdoings deserve excuse.

An amendment is needed to make this condition on blameworthiness minimally plausible. I started with the assumption that Vican and Erdemović are fully virtuous. But this is not obvious. It was never part of their description that these two agents were fully virtuous people. The following amendment may help: in order to warrant excuse for wrongdoing, one must be fully virtuous *in the relevant respect*. If there is more than one virtue, each one will have a different domain of relevance. When someone performs a wrongdoing, they skirt a requirement that will be in the domain of one of the identified virtues. If someone is excused for their wrongdoing, they must be fully virtuous in the relevant way, i.e. they must have the relevant virtue. Is it clear that Vican and Erdemović are fully virtuous in the relevant respect? That is, do they have the virtue that protects against the kind of requirement they skirt? It may be hard to tell without adopting a substantive account of the virtues. But plausibly both agents have the disposition to refrain from harming others, or the disposition to protect life, or perhaps the disposition to avoid causing pain to others. The psychological turmoil they experience is evidence of their virtuousness in this respect. This is why these agents deserve an excuse.

5. Conclusion

Full virtue can get someone out of trouble. That is, being fully virtuous will block blameworthiness even in the face of wrongdoing. But, unlike the psychological pressure views discussed earlier, there is no concern that this excuse can be abused. Being virtuous is hard work. A fully virtuous person has the beliefs, desires, hopes, aversions, and knowledge that characterize virtue. Since these states ground a disposition to behave well, it will be rare—virtually never in ordinary circumstances—that a fully virtuous person will make use of an excuse. But in the event that her circumstances exert the kind of pressure that is typical of duress cases, she may be afforded some excuse for her wrongdoing. In these cases, her wrongdoing is excused because it does not extinguish or diminish her virtuous disposition.

One final advantage of my proposed explanation deserves special mention. If we adopt any norm-splitting explanation of excuse by duress, we are faced with an open question: why is morality fragmented in the way required by norm-splitting explanations? Unless a rationale can be given, it might appear that deviations in moral norms are merely pre-theoretical confusions that must be resolved by philosophical reflection.⁴³ I suspect that this open question will be difficult to answer in any principled way. To its credit, the proposed psychological explanation does not generate this troubling question. On the view, morality speaks univocally; it tells us to behave in particular ways, care about acting in those ways, and cultivate dispositions to act accordingly. Cases of duress tell us that there are discrepancies in our judgments about wrongfulness and blameworthiness, but these cases needn't push us to abandon a unified conception of moral norms. A simple distinction between full and infallible virtue gives us the

⁴³ Rosen (2014: 86) acknowledges that his norm-splitting proposal generates this question. This appears to be one reason he abandons norm-splitting and provides an explanation grounded in 'Scanlon-style social distancing.'

resources to capture these discrepancies in judgment without attributing any peculiar doubletalk to morality.

CHAPTER THREE

Blameless Badness

One ordinary sense of responsibility corresponds to blameworthiness.¹ To say that an agent is responsible in this sense is to say that it is appropriate to direct the reactive attitudes, e.g. resentment and indignation, toward him in response to his behaviour. In a classic essay, Gary Watson identifies a sense of responsibility that he claims doesn't amount to the judgment that a person is blameworthy.² Consider someone who sabotages herself by drinking heavily on the night before an important performance. Watson claims there is a sense in which she is responsible for her late-night drinking, but it is not appropriate to direct resentment or indignation toward her for this behaviour. He claims she is responsible in the *attributability* sense even though she is not blameworthy.

Watson's view has taken on the status of dogma in the study of responsibility.³ Most discussions of responsibility presuppose the distinction between blameworthiness and attributability. The notion of attributability has become the subject of investigation in its own right, though substantive debate about it appears to have reached an impasse.⁴ Despite its pervasive appeal, I believe that the concept rests on a confusion. The aim of this chapter is to bring out this confusion, and to argue that a second sense of responsibility is not needed in any

¹ I focus in this chapter on attributability in contrast to blameworthiness, leaving the discussion of praiseworthiness until the last section.

² Watson (1996).

³ There are detractors, though. See Smith (2008b, 2012, 2015) and Wolf (2015)

⁴ I address this point in §3.

case. The account of blameworthiness described in Chapter Two can make sense of the cases that worry philosophers like Watson.

What is attributability? Watson begins by establishing that attributability does not merely identify the agent as having a “causal role...of special importance to the explanation of [the action]” (Wolf 1990: 40 cited in Watson 1996: 232). He is especially keen to avoid this ‘shallow’ description of attributability; he means to pick out a genuine kind of *moral* appraisal. For him, to say that an action is attributable is to say that the relationship between the agent and her action is “not (just) causal, but *executive* and *expressive*” (Watson 1996: 233 his emphasis). What exactly does the action express? An attributable action expresses “what I’m about, my identity as an agent” (1996: 233). It “[concerns] the agent’s excellences and faults – or virtues and vices – as manifested in thought and action” (1996: 231). Some of Watson’s other remarks are now associated with a substantive view about attributability. But the concept he described has been adopted widely. David Shoemaker, for example, explains attributability judgments thusly:

[They are] responses to revealed traits of *character*. When I disdain someone in light of his pattern of psychological spousal abuse, I am, as Watson notes, tracing these attitudes and actions to a fault, a vice, of his. By contrast, when an otherwise placid, loving person who has become involuntarily intoxicated snaps and yells at her spouse, we tend not to disdain her (or admire her, for that matter), that is, we tend not to attribute such an outburst to her for purposes of responsibility, as we view it to have been genuinely *out of character*. It just did not express who she really is (Shoemaker 2015: 118)

Many other philosophers describe the notion along similar lines, though they sometimes swap ‘character’ for ‘deep self’ or ‘moral identity.’ For example: “attributable actions specifically express the agent’s moral identity and so open him up to distinctive forms of moral evaluation – namely, evaluations of him as a moral agent” (Strabbing 2016: 290). When an action is attributable, we “need only judge that the conduct is faulty in some way that reflects poorly on the agent” (Fisher & Tognazzini 2011: 3). Something that “is attributable to the agent is either a *reflection* or an *expression* of the agent’s self” (Driver 2015: 165 her emphasis). I will for the most part ignore the difference between ethical character, moral identity, and the self. These terms are all being used as names for a “privileged subset of our psyche” that comprise who we are, ethically-speaking (Shoemaker 2015: 118).

This chapter raises trouble for attributability in three stages. First, I point out an internal tension in the concept. As usually understood, it cannot make sense of the attributability of our virtues and vices, and as far as I can see, these appear to be candidates for attributability as much as anything else. Second, I offer a hypothesis about the source of this tension, which also explains why substantive debate about attributability appears to have stalled. Third, I examine the cases that motivate the concept in the first place, ultimately finding that my earlier theory of blameworthiness can explain the cases that prompt the turn to attributability.

1. Motivating Attributability

The concept of attributability amounts to the following: to say that X is attributable to an agent is to say that it reflects or expresses the agent’s ethical character.

There are two preliminary points to make about the concept under investigation. First,

although I will often focus on attributable actions, a person's attitudes, feelings, and omissions can be attributed to her.⁵ Second, I do not mean to give a substantive view about attributability. When giving a substantive view, one aims to give necessary and sufficient conditions for the concept's application in terms that do not deploy the concept. The description of attributability above is much more modest; it aims to paraphrase the concept in terms that can be agreed on by all parties involved in substantive debate. In this chapter, I mean to criticize the concept itself, not any particular way of analyzing it.

Of course, it is not always fair to dismiss a concept because it is hard to paraphrase or untangle. It is not as though we must always settle on a paraphrase before engaging in philosophical inquiry about a concept. When studying 'folk' concepts, which appear to us independently of our theoretical ambitions, it may be fair to leave the paraphrasing to one side. But the notion of attributability was identified through philosophical reflection, so it is reasonable to expect that a working interpretation of the concept be given. This is especially important when substantive debate about the concept has stalled.

Now, what motivates the concept of attributability in the first place? As I mentioned, Watson notices that there appears to be a sense in which a person can be responsible for a behaviour even though she is not blameworthy for having performed it. Someone who "endangers something of deep importance to her life for trivial ends" by "sleeping too little and drinking too much before important performances" is not blameworthy for her conduct (Watson 1996: 231).. After all, unless I believe she has some bizarre obligation to me that she achieve her ends, it would make little sense for me to respond to her with resentment or indignation. Still, Watson insists, there is no doubt this person has acted imprudently or unwisely—*badly* in short.

⁵ On this point, see Smith (2005).

This action “expresses the agent's own evaluative commitments, her adoption of some ends among others” (Watson 1996: 233). Since adopting an end “is to declare what one stands for,” he thinks we have identified a sense of responsibility distinct from blameworthiness (Watson 1996: 233).

Harry Frankfurt provides another familiar case used to motivate the concept of attributability.⁶ Frankfurt compares two addicts. The first addict is *unwilling*. He “tries everything that he thinks might enable him to overcome his desires for the drug” (Frankfurt 1971: 12). Despite his efforts, his desires are too powerful to resist, and he ends up indulging his addiction. The second addict is *willing*. He fully embraces his addiction and indulges it without resistance or displeasure. Even though both addicts act intentionally, the willing addict appears responsible for his drug-taking in a way that the unwilling addict is not. The concept of attributability is supposed to explain the difference. While the willing addict’s behaviour is attributable to him, the unwilling addict’s behaviour is not. Taking the drug reflects the willing addict’s ethical character; he is an imprudent or self-indulgent person and his drug-taking expresses this vice. But taking the drug does not seem to reflect the unwilling addict’s ethical character. Given his efforts to resist the drug, he appears prudent and not at all self-indulgent, but his action fails to reflect this. It fails to reflect who he really is, ethically-speaking.

In the light of these examples, the concept of attributability can appear indispensable. Similar cases involve agents who act under hypnosis, suffer from mental illness, or make simple mistakes. Still, I think the concept of attributability is ultimately confused. The next section introduces my main challenge for the concept.

⁶ See Frankfurt (1971).

2. Internal Tension

The basic problem for attributability is that our virtues and vices appear to be a real candidates for attributability, just as much as our actions, attitudes, and feelings are. But, since the fact that a thing reflects or expresses an agent's ethical character makes it attributable to the agent, it is either unintelligible or trivial to say that an agent's virtues and vices are attributable to her. This points to an internal tension in the concept.

Let me first focus on the case of vice. Often, one does not come to have a character trait spontaneously. Perhaps this is conceptually impossible, because traits require long-lasting and sustained dispositions, or perhaps it is merely unlikely, because of the way our psychological dispositions happen to change. In either case, it is possible that someone might find out he has a vice, and in the process of improving himself, truly disavow the vice such that it is no longer attributable to him. Consider the following case.

A childhood friend re-enters my life, and tells me that I have become rather self-involved. I tend to steer conversations so that they concern me, disregard the interests of others in making decisions, and forget to put my own tribulations in perspective. My friend reminds me that I was not always so self-involved, and on her urging, I resolve to change my ways. I take every effort to make the change feasible. I ask my colleagues to let me know if I change the subject of conversation, I learn more about the perspectives of others by reading widely, and I set up an alert to remind me to consult others before making decisions. In the early stages of my development, before my efforts have paid off, one might still say I am self-involved. But is this trait attributable to me? For someone convinced by the earlier examples, it is not clear why it would be. I have tried everything I can to avoid this vice's manifestation, and committed myself

to its eradication, even though it still manages to find expression in me. In this regard, I am like the unwilling addict, who does not embrace his addiction or its manifestations and takes every effort to resist acting as he does.

One feature of Frankfurt's case is that the unwilling addict appears to be a helpless victim of his compulsion. His 'deep self' is sidelined while his addiction runs rampant. There is certainly something strange about the idea that someone can be a victim of her own vice, or that her vice can run rampant while she is inactive. But consider the possibility that a person can be a victim, so to speak, of her virtue. Imagine a kind of 'dirty hands' case: in order to achieve the end of saving lives, it is imperative that an army surgeon become largely insensitive to the pain of the soldiers whose limbs he must amputate.⁷ The surgeon's job is to get through the amputation, and sensitivity to the patient's pain will interfere with this task. An aspiring but compassionate surgeon will need to cultivate this rather callous disposition by taking every effort to resist her compassion. In the early days of training, she has a trait of character that she deeply wants to diminish, and judges that she ought to diminish all-things-considered. In this case, it is quite natural to see her compassion as an interference with the expression of her 'deep self' just as the addict's compulsion interfered with the expression of his 'deep self.' It is at least very difficult to point to an asymmetry in the cases.

If my self-involvement and the surgeon's compassion are not attributable, then it is not trivial or unintelligible to attribute parts of someone's ethical character to that person. But the concept makes the attributability of one's ethical character unintelligible or trivial, and thus it appears attributability faces a kind of internal tension; the concept seems to apply in ways that are unintelligible or trivial by the lights of the concept itself.

⁷ For more on 'dirty hands' cases, see Walzer (1973), Williams (1978), and Nielsen (2007).

There are two ways to resist the claims above. The first is to deny that the disavowed virtues and vices are truly parts of the agent's 'deep' self—the self that matters for attributability. The second is to claim that someone cannot manifest a character trait inactively; if I am always 'active' in manifesting my traits, then I cannot disavow them as the examples above propose. I address these in turn.

The first way around my challenge seems futile. In the argument above, I identified a person's ethical character with the person's virtues and vices. If ethical character is supposed to be 'deeper' than one's virtues and vices, what could it be? Perhaps we can say that one's ethical character is constituted by one's *deep*-virtues and *deep*-vices. To say that X is attributable to an agent is to say that X reflects or expresses the agent's deep-virtues and deep vices. But it seems to me that similar cases can be generated, at least until more is said about the idea of deep-virtues and deep-vices. Suppose someone identifies one of his deep-vices, and judges that he ought to change it, cares deeply about changing it, and makes policies in order to change it. It seems that the deep-vice will no longer be attributable. But this possibility is either trivial or unintelligible on the reading of attributability above, so the internal tension reappears.

The second response is more promising. In a recent essay, Susan Wolf makes the claim that character traits cannot be exercised inactively, although she is addressing a different problem.⁸ Focusing on the difference between traits of character and disorders, she proposes the following: in contrast to the "stiffness" of manifesting a disorder, exercising a virtue or vice requires *active intelligence* (Wolf 2015: 370). The capacities involved in active intelligence require "a flexibility and responsiveness...the realization of which depends on the ongoing exercise of powers of perception, imagination, reason, empathy, and more" (Wolf 2015: 368).

⁸ Wolf offers this proposal tentatively. See her comments in Wolf (2015: 368).

She goes on to draw the implications of this thought for responsibility. She says:

A person is *necessarily* active when her character is engaged—she is necessarily exercising powers of perception, imagination, reasoning, and so on—monitoring and assessing innumerable aspects of her situation and of the values that bear on the question of how to respond to it (Wolf 2015: 369).

[] I do think that the fact that we have and exercise such powers is relevant to and perhaps even at the heart of our status as responsible beings. It is only in virtue of being able to have and to exercise such powers, I think, that we can be responsible for anything at all (Wolf 2015: 369).

Wolf's idea is that ethical traits *ground* the agent's status as the sort of creature for whom things can be attributable. So, if one accepts her view, the virtues and vices may not themselves be real candidates for attributability. In effect, Wolf has provided the crucial asymmetry between traits of character and the other things commonly attributed to agents, like actions, attitudes, and feelings. While it is possible to act, think, or feel without exercising active intelligence, the expression of virtues and vices *require* the exercise of active intelligence.

Wolf's discussion of the relationship between character and responsibility is both novel and interesting. But I doubt her account of the virtues and vices. My objection is that some character traits are better understood not as the exercise of active intelligence—capacities of perception, imagination, and reasoning—but as *the failure* to exercise these capacities. Consider modesty. Although the precise nature of modesty is controversial, it appears to be naturally

described as failing to exercise an ‘active’ capacity. Julia Driver, for example, has argued that modesty involves epistemic defect: “for a person to be modest, she must be ignorant with regard to her own self-worth” (1989: 377). It is hard to imagine that this kind of ignorance results from exercising capacities of perception, imagination, or reasoning. After all, if these capacities are exercised, one is unlikely to underestimate one’s self worth as required for modesty. Other views of modesty make a similar point. Nicholas Bommarito has recently argued that modesty “involves certain patterns of conscious attention, which are characterized by an *inattentiveness* to good qualities that reflect well on oneself, the value of such qualities, and one’s own role in bringing them about” (2013: 111 his emphasis). If modesty is any indication, it seems a person is not always active in Wolf’s sense when exercising a trait of character.

Wolf considers the case of dishonesty, admitting that it seems naturally described as an incapacity, but insisting that “the assessment of a situation as one in which lying *will* help one achieve one’s goals, as one that, from the perspective of the dishonest person, calls for lying involves the same kind of flexibility and sensitivity to nuance and detail that is needed for the possession and display of honesty” (Wolf 2015: 368). But it seems that someone can count as fully dishonest without possessing any of the capacities Wolf mentions, as long as she totally and completely lacks the capacities required for honesty. This person may not lie in order to achieve any goals; she may lie simply because she places no value on the truth whatsoever and never perceives a situation as calling for honesty. If the absence of virtue is sufficient for vice, exercising a vice needn’t always involve active intelligence, either.

But could someone be a responsible agent without having *any* ‘active’ traits whatsoever? The answer is probably no. But this is irrelevant to the task at hand. Even if we grant that an agent must have *some* constellation of active traits in order to be a responsible agent, it still

seems intelligible and non-trivial to ask whether a *given* trait—say, her modesty or dishonesty—is attributable. And, as long as this question makes sense, the argument against attributability proceeds. Individual virtues and vices appear to be real candidates for attributability, but by the lights of the concept, this should be unintelligible or trivial.

3. A Divided Concept

The argument in the previous section may appear somewhat thin. Doesn't it simply show that we must characterize attributability without appealing to ethical character or the deep self? Isn't that where the problem lies? Indeed, there are ways of understanding attributability that do not explicitly appeal to the deep self or ethical character. Nomy Arpaly and Timothy Schroeder have suggested that the attributability of an action has to do with the extent to which the psychological features underlying the action are integrated in the agent's personality.⁹ In this section I want to point out the more general conflict in the concept that will seal my case against it, regardless of how we choose to spell it out.

It seems to me that the concept of attributability straddles two distinct projects. The first project is generated by the apparent possibility of acting *badly* without being *blameworthy*.¹⁰ It is possible, Watson noted, to judge that someone has acted unwisely, foolishly, or imprudently—*badly* in short—without judging that they are blameworthy. The self-sabotaging agent, who sacrifices something she values tremendously for something trivial, acts badly but is purportedly not blameworthy. This generates a project: what does it mean to act badly, and how does it

⁹ See Arpaly and Schroeder (1999)

¹⁰ See Levy (2005). I set aside the possibility of acting well but not being praiseworthy. Watson seems to want to leave this open, but he doesn't provide a compelling case for it. I discuss this further in §5.

contrast with being blameworthy for an act?

The second project concerns self-identification. People who suffer from an addiction or compulsion may desire to act as their addiction or compulsion requires, but this desire is somehow *not their own*. Likewise, in cases of depression or anxiety, a person may be alienated from their thoughts, feelings, and behaviours. The interference of fear, sadness, anger, and sexual drive may cause someone to act in ways that they experience as foreign.¹¹ Thus we have a second project: which parts of me count as *mine*?

A proponent of attributability seems to think the ‘self’ pursued in the second project has something to do with the judgments of acting badly in the first project. When we judge that someone acts badly (as opposed to blameworthily), we judge that their action reflects a deficiency in the parts of her psychology that she identifies as her own.¹² But it is not straightforward why these projects should be connected in this way. They certainly do not appear to bear any *necessary* connection to one another; their marriage is one of mere convenience.

And it may not be a particularly happy marriage. On one hand, a plausible answer to the first project seems like an inadequate answer to the second. It may be plausible that acting badly involves the expression of one’s character, but it is not plausible that a person always identifies with every aspect of ethical character. This was the point made in the previous section. On the other hand, a plausible answer to the second project seems like a lousy answer to the first. It may be that I identify most with the actions that issue from fully integrated features of my psychology, as Arpaly and Schroeder suggest; but it is unclear why acting badly has anything to

¹¹ For an overview of the literature on attributability, see Shoemaker (2015).

¹² I will omit the phrase ‘as opposed to judging that they are blameworthy’ going forward. The topic will continue to be the nature of acting badly, as contrast with blameworthiness.

do with integration. David Shoemaker raises precisely this concern against Arpaly and Schroeder. He asks: “But what exactly does *integration* have to do with our [aretaic] praise and blame?” (2015: 137 his emphasis). He continues: “some attitudes could just be well-integrated out of long-standing habit or laziness, and it is difficult to see their integration in that case as the source of any aretaic praise or blame we might muster” (Shoemaker 2015: 137).

Consider the substantive debate about attributability. Though philosophers have located the self in many different parts of our psychology, proposals divide into two camps: cognitive and non-cognitive views. The cognitive approach locates the self in states like evaluative judgments and commitments, while the non-cognitive approach locates the self in states like cares or desires.¹³ The latter camp is unable to accommodate special cases where a person’s cares come apart from her judgments. A depressed person may not care about her health, but she may well judge that this attitude is unjustified and commit to changing it. In doing so, she seems to distance herself from her self-destructive actions; they no longer appear attributable to her. The former camp is unable to accommodate reverse cases. A parent may care for her delinquent son even though she judges that helping him is the wrong thing to do.¹⁴ The parent’s acts may be hers even though they conflict with her considered judgment. It has proved so difficult to choose either path that the most promising approaches to the concept are hybrid or ecumenical.¹⁵

The intractability of this debate may be partly explained by the diagnosis above. It may be that acting badly (in the sense above) has nothing to do with the parts of one’s psychology with which one identifies. Since the literature on attributability has presupposed this connection,

¹³ See Shoemaker (2015) for an overview. For examples of the cognitive approach, see Watson (1975) and Smith (2008b, 2012, 2015). See Frankfurt (1971), Shoemaker (2003), Jaworska (2007) and Sripada (2015) for examples of the latter.

¹⁴ This case comes from Shoemaker (2011: 611).

¹⁵ See Bratman (1999: 185–206) for a hybrid view. Shoemaker (2015) offers an ecumenical alternative.

no single view will accommodate everything subsumed under the concept.

4. Acting Badly

I started my critique by offering an objection to the concept of attributability; it cannot make sense of the attributability of our virtues and vices, and as far as I can see, these appear to be candidates for attributability as much as anything else. I extended this critique by offering a more general characterization of the problem; the concept presupposes there is a connection between two projects that needn't bear any connection to one another. In this final stage of critique, I want to address the motivation behind the concept in the first place. Do we need a concept of responsibility beyond blameworthiness to explain the difference between acting badly and being blameworthy? Or to explain cases like Frankfurt's unwilling addict? In this section, I argue that the answer is no. The cases of interest can be accommodated with the proper theory of blameworthiness.

Recall the kind of case that motivates Watson. In one example, he imagines someone who "endangers something of deep importance to her life for trivial ends" by "sleeping too little and drinking too much before important performances" (1996: 231). Watson emphasizes that other people do not have *standing* to blame this agent; it is 'nobody's business' that she has made this imprudent decision. But this may not go far enough. After all, given the importance she places on her performance, it would certainly be appropriate for the agent to blame *herself* for her conduct. In their first-personal guise, blaming attitudes usually take the form of remorse, regret, self-reproach, or guilt.¹⁶ Watson explicitly puts the 'self-reflexive' case to one side,

¹⁶ This idea first appears in Williams (1985:177).

though it is not clear why he does so. Like resentment directed toward others, the attitudes of self-reproach, guilt, and remorse “involve a readiness to adverse treatment” and can be appropriate in some circumstances and not others (Watson 1996: 239). Someone who feels guilt or remorse for something is likely to deprive herself of pleasures, feel anger and hatred toward herself, and take measures to ensure she never slips up in that way again.

It is not negligible that blaming attitudes are appropriate in Watson’s cases, even in this limited way. If this is granted, it can seem like attributability is the mere byproduct of blameworthiness. To judge that an action is attributable is to acknowledge that I lack standing to blame but recognize that someone else, or the agent herself, could justifiably react with a readiness for adverse treatment. On such a view, the judgment of attributability may just be a belief *about* blameworthiness. It would hardly qualify as a sense of responsibility distinct from blameworthiness.

It is merely incidental, then, that the agent in Watson’s example can justifiably respond to her act with attitudes like remorse, guilt, or self-reproach. He intends to give a case where someone is assessable for what they do, but *no one* can justifiably respond with blaming attitudes. Not me, or some other bystander, or the agent herself. How does such a case look? Suppose I promise my sick friend that I will have his body cremated after his death and scatter his ashes in a distant and remote part of the ocean. But after his cremation, the voyage is repeatedly canceled because of bad weather and problems with the travel company. I eventually forget about the ashes. A few months later, I notice the ashes again and decide to put them in my attic, effectively deciding that they will not be scattered as my friend desired. In this case, there may be no one who can justifiably blame or sanction me for my promise-breaking; I didn’t break a promise to anyone but my deceased friend. Although I may feel some guilt for what I have

done, this feeling may not be deserved; I promised in good faith and made a reasonable effort to do what I promised. But it appears that I have acted badly, even though I should not be targeted by resentment, indignation, or disapproval.

In what sense do I act badly? The proponent of attributability may say that my action expresses a deficiency in the subset of myself with which I identify as an agent; it expresses apathy, lack of concern, or disloyalty. But this isn't the only way to understand the badness of my act. To see why, recall the second chapter of this dissertation.

My thought in Chapter Two was that we shouldn't think the only interesting kind of virtue is perfect compliance with morality. Instead, we should distinguish two importantly different states: full virtue and infallible virtue. Full virtue requires that one have the disposition to act well, grounded in the beliefs, desires, hopes, aversions, and other attitudes that characterize virtue.¹⁷ Infallible virtue requires perfect compliance with morality in cases where one is not ignorant about relevant matters of fact. I used this distinction to articulate a principle about blameworthiness: fully blameworthy acts are those incompatible with the agent's possession of full virtue, i.e. the disposition to act well.

The distinction between full and infallible virtue allows us to distinguish the acting badly and blameworthiness in the following way. If the wrongdoing speaks against the agent's possession of full virtue, she is *blameworthy* for performing it. If the wrongdoing merely speaks against her infallible virtue, she has *acted badly*.¹⁸

To see the account in action, recall the central cases in Chapter Two. Imagine a young soldier who kills an innocent group of people while under genuine duress, e.g. while threatened

¹⁷ One can be fully virtuous *in a respect*. To be fully virtuous in a respect is to have a more limited disposition to act well, grounded in the attitudes that characterize the particular virtue.

¹⁸ Notice that, given the definition of infallible virtue, wrong actions performed in ignorance do not count against one's infallible virtue.

with death if he doesn't comply with the order to kill the innocents. Though the details matter, I claimed it is possible to give a version of this scenario where the agent is genuinely *excused*; he acts badly, but he is not blameworthy.¹⁹ I offered the theory above to explain this possibility: the soldier's behaviour merely counts against his infallible virtue, but is compatible with his possession of full virtue. His behaviour is bad, but he is not blameworthy.

The present case is amenable to similar treatment. It doesn't seem appropriate to blame or sanction me for breaking my promise to my deceased friend, but I act badly in failing to do what I promised. My good faith attempt to fulfill my promise is evidence of my full virtue, or disposition to act well, with respect to promise-keeping. But my disposition tolerates some failures to manifest, i.e. some promise-breakings. Breaking my promise to a deceased friend is one such case. Breaking this promise merely speaks against my infallible virtue, so it is not blameworthy even though it is bad.

How do we decide which acts merely count against infallible virtue, and which ones count against the possession of full virtue? This task is the joint work of moral theory and folk psychology.²⁰ We first need a moral theory that specifies the virtues in all their psychological richness—in terms of the beliefs, desires, hopes, feelings and other states that ground dispositions to act well. With that account in hand, it is up to folk psychology to tell us what kinds of behaviour are consistent with the maintenance of those attitudes and the dispositions they ground.

Though specific examples may be controversial, at least some kinds of bad behaviour are compatible with the possession of full virtue. Wrongdoings performed in ignorance are certainly

¹⁹ See Chapter Two, §1.

²⁰ See Chapter Two, §3.

compatible with full virtue.²¹ If I do not know that my friend is mourning the death of a family member, I may send her an invitation to a party with a grim or dark theme. This is a bad thing to do; it is insensitive to my friend's pain. But it seems to me that I may nonetheless possess all of the attitudes that characterize sensitivity. Wrongful behaviour in ignorance is compatible with my possession of the relevant dispositions and the attitudes that ground it.

Further, people in the grips of anxiety often find it difficult to focus and may consequently fail to pay attention to colleagues who seek personal advice from them. This too is a bad thing to do; it is discourteous and a bit callous. But anxiety appears compatible with possessing the attitudes and dispositions that constitute courtesy and kindness. We do not describe anxious people as self-involved or narcissistic. The attitudes required for narcissism may indeed be incompatible with the relevant dispositions to act well, but a proclivity for episodes of anxiety is not.

Consider Frankfurt's unwilling addict once more. On my view, the unwilling addict acts badly, but is not blameworthy. This is because he possesses the disposition to act well, as evidenced by his efforts to resist the drug, even though he ultimately indulges the addiction. His drug-taking behaviour merely speaks against his infallible virtue. But the willing addict has no such disposition to act well; he has no desires, beliefs, hopes, or interests that dispose him to avoid the drug. Thus he is blameworthy for what he does.

The concept of attributability has made the explanation of excuse far more difficult than it needs to be. If we accept attributability, the explanation of the unwilling addict's excuse proceeds like this. The unwilling addict's behavior is not attributable, i.e. this behaviour doesn't reflect his deep self, and since blameworthiness requires attributability, the unwilling addict is

²¹ They are also compatible with infallible virtue, as stipulated in the definition.

not blameworthy. As alluring as this explanation may be, it makes other cases of excuse deeply puzzling. What about actions we perform under duress or in self-defense? On one hand, if acts performed under duress or in self defense are attributable, then we need some additional story about why they are not blameworthy. On the other hand, if acts performed under duress or in self defense are not attributable, then it is puzzling why they have acted badly. After all, if their behaviour is not attributable, they have not acted badly, according to the proponent of attributability. My proposal cuts out attributability altogether. I give an account of blameworthiness in terms of the expression of character, which I think is capable of explaining all cases of bad conduct for which the agent is not blameworthy.

5. Conclusion

My goal in this essay has been to argue that the concept of attributability should be jettisoned from the philosophy of responsibility. I argued that there is an internal tension in the concept (§2). I claimed that the tension is owed to the fact that attributability connects acting badly with the expression of one's self-identity, i.e. the parts of oneself with which one identifies (§3). Further, I showed that the cases used to motivate the concept of attributability can be explained by a theory of blameworthiness, and that attributability muddies the explanation of excuse (§4). I want to end by addressing one final point.

There is an asymmetry between praiseworthiness and blameworthiness on my view that is not present in the concept of attributability. The view advanced in this dissertation does not distinguish acting well and praiseworthiness, but it does distinguish acting badly and blameworthiness. But, as Watson presents it, attributability is open to application in cases of

good and bad conduct alike. He suggests that we can not only distinguish acting badly and blameworthiness, but we can also distinguish acting well and praiseworthiness. One might say: acting well is a matter of expressing an *excellence* of ethical character, while praiseworthiness is a matter of the appropriateness of directing *positive* attitudes toward the agent. Thus there is an asymmetry in my view, but not the concept of attributability.

An asymmetry here seems right.²² There may be no deep sense in which acting well comes apart from praiseworthiness, even though there *is* a difference between acting badly and being blameworthy. Of course, I may judge that a person conducts herself well—say, by holding the door for me—but not respond by “bestowing a medal or...remarking on the person's merits” (Watson 1996: 242). But this doesn't clearly establish that acting well and praiseworthiness come apart. The example shows that a judgment of full praiseworthiness leaves an open question, namely: how do I express my attitudes in action? In cases like holding the door, I might nod in appreciation or smile, but in cases of heroism, I may perform more elaborate acts of appreciation. The same question is open after judgments of blameworthiness. When someone deliberately slams the door in my face, and I judge that she is fully blameworthy, the same question remains: how should I express these attitudes in action? For a small wrongdoing like shutting the door, a sneer or cold-shoulder may suffice, but more serious cases call for interventions like imprisoning, punishing, and ostracizing.

In this chapter, I have tried to dislodge the concept of attributability from the philosophy of responsibility. The concept has led us to an intractable debate about where our real selves dwell among the things that make us who we are. My hostility toward attributability was predictable from the account in Chapter Two. I have generally tried to show that ethical character

²² Wolf (1980) discusses this asymmetry between praise and blame.

is more intimately tied to responsibility than the concept of attributability would allow. If given the chance, the virtues and vices can offer a much more powerful, fruitful, and systematic explanation of responsibility.

REFERENCES

- Allen, Bem P., and Charles R. Potkay. 1981. "On the Arbitrary Distinction between States and Traits." *Journal of Personality and Social Psychology* 41 (5): 916–28. doi:10.1037/0022-3514.41.5.916.
- Annas, Julia. 2004. "Being Virtuous and Doing the Right Thing." *Proceedings and Addresses of the American Philosophical Association* 78 (2): 61–75.
- . 2005. "Comments on John Doris's Lack of Character." *Philosophy and Phenomenological Research* 71 (3): 636–42.
- . 2011. *Intelligent Virtue*. OUP Oxford.
- Armstrong, David M. 1961. *Perception And The Physical World*. Humanities Press.
- Arpaly, Nomy. 2002. "Moral Worth." *Journal of Philosophy* 99 (5): 223–45.
- . 2003. *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford University Press.
- Arpaly, Nomy, and Timothy Schroeder. 1999. "Praise, Blame and the Whole Self." *Philosophical Studies* 93 (2): 161–88.
- . 2014. *In Praise of Desire*. OUP.
- Ashwell, Lauren. 2010. "Superficial Dispositionalism." *Australasian Journal of Philosophy* 88 (4): 635–53.
- . 2015. "Conflicts of Desire: Dispositions and the Metaphysics of Mind." In *Putting Powers to Work*, edited by Jonathan Jacobs. Oxford: Oxford University Press.
- Attfield, David G. 1978. "Motivation in Moral Education: The Case for Virtue." *Journal of Moral Education* 7 (3): 158–65.
- Austin, John. 1956. "A Plea for Excuses." *Proceedings of the Aristotelian Society* 57: 1–30.
- Baron, Marcia. 1984. "The Alleged Moral Repugnance of Acting From Duty." *Journal of Philosophy* 81 (4): 197–220.
- . 2006. "Excuses, Excuses." *Criminal Law and Philosophy* 1 (1): 21–39.
- . 2014. "II—Culpability, Excuse, and the 'Ill Will' Condition." *Aristotelian Society Supplementary Volume* 88 (1): 91–109.

- Bayles, Michael D. 1982. "Character, Purpose, and Criminal Responsibility." *Law and Philosophy* 1 (1): 5–20.
- Bennett, Jonathan. 1979. "Accountability." In *Philosophical Subjects: Essays Presented to P.F. Strawson*, edited by ZV Straaten. Oxford: Oxford University Press.
- Bird, A., and T. Handfield. 2008. "Dispositions, Rules and Finks." *Philosophical Studies* 140 (2): 285–98.
- Bird, Alexander. 1998. "Dispositions and Antidotes." *Philosophical Quarterly* 48 (191): 227–34.
- . 2007. *Nature's Metaphysics: Laws and Properties*. Oxford University Press.
- Bommarito, Nicolas. 2013. "Modesty as a Virtue of Attention." *Philosophical Review* 122 (1): 93–117.
- Brandt, R. B. 1958. "Blameworthiness and Obligation." In *Essays in Moral Philosophy*, edited by A. I. Melden. University of Washington Press.
- Brandt, Richard B. 1970. "Traits of Character: A Conceptual Analysis." *American Philosophical Quarterly* 7 (1): 23–37.
- Bratman, Michael. 1999. *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge University Press.
- Brooks, Rosa E. 2012. "Law in the Heart of Darkness: Atrocity and Duress." *Virginia Journal of International Law* 43 (February).
- Chaplin, W. F., O. P. John, and L. R. Goldberg. 1988. "Conceptions of States and Traits: Dimensional Attributes with Ideals as Prototypes." *Journal of Personality and Social Psychology* 54 (4): 541–57.
- Choi, S. 2012. "What Is a Dispositional Masker?" *Mind* 120 (480): 1159–71.
- Choi, Sungho. 2005. "Do Categorical Ascriptions Entail Counterfactual Conditionals?" *Philosophical Quarterly* 55 (220): 495–503.
- . 2009. "The Conditional Analysis of Dispositions and the Intrinsic Dispositions Thesis." *Philosophy and Phenomenological Research* 78 (3): 568–90.
- Clarke, Randolph. 2008. "Intrinsic Finks." *Philosophical Quarterly* 58 (232): 512–18.
- . 2010. "Opposing Powers." *Philosophical Studies* 149 (2): 153–60.

- Cokelet, Bradford. 2015. "Dispositions, Character, and the Value of Acts." In *Character: New Directions from Philosophy, Psychology, and Theology*, edited by Christian Miller, R. Michael Furr, Angela Knobel, and William Fleeson. Oxford University Press.
- Crisp, Roger. 2010. "Virtue Ethics and Virtue Epistemology." *Metaphilosophy* 41 (1): 22–40.
- . 2015. "A Third Method of Ethics?" *Philosophy and Phenomenological Research* 90 (2): 257–73.
- Dancy, Jonathan. 1993. *Moral Reasons*. Blackwell.
- . 2004. *Ethics Without Principles*. Oxford University Press.
- Davidson, Donald. 1963. "Actions, Reasons, and Causes." *Journal of Philosophy* 60 (23): 685–700.
- Doris, John M. 2002. *Lack of Character: Personality and Moral Behavior*. Cambridge University Press.
- . 2005. "Replies: Evidence and Sensibility." *Philosophy and Phenomenological Research* 71 (3): 656–77.
- Driver, Julia. 1989. "The Virtues of Ignorance." *Journal of Philosophy* 86 (7): 373–84.
- . 2001. *Uneasy Virtue*. Cambridge University Press.
- . 2012. "Luck and Fortune in Moral Evaluation." In *Contrastivism in Philosophy: New Perspectives*, edited by Martijn Blaauw. Routledge.
- . 2015. "Appraisability, Attributability, and Moral Agency." In *The Nature of Moral Responsibility*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith, 157–74. Oxford University Press.
- Duff, Antony. 2007. *Answering for Crime: Responsibility and Liability in the Criminal Law*. Hart Pub.
- Duff, R. A. 2006. "Excuses, Moral and Legal: A Comment on Marcia Baron's 'Excuses, Excuses.'" *Criminal Law and Philosophy* 1 (1): 49–55.
- Fara, Michael. 2005. "Dispositions and Habituals." *Noûs* 39 (1): 43–82.
- . 2008. "Masked Abilities and Compatibilism." *Mind* 117 (468): 843–65.
- Finkelstein, Claire. 2002. "Excuses and Dispositions in Criminal Law." *Buffalo Criminal Law Review* 6 (1): 317–59.

- Fischer, John Martin, and Neal A. Tognazzini. 2011. "The Physiognomy of Responsibility." *Philosophy and Phenomenological Research* 82 (2): 381–417.
- Foot, Philippa. 1985. "Utilitarianism and the Virtues." *Mind* 94 (374): 196–209.
- Frankfurt, Harry G. 1969. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66 (3): 829–39.
- . 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68 (1): 5–20.
- Fridhandler, Bram M. 1986. "Conceptual Note on State, Trait, and the State–trait Distinction." *Journal of Personality and Social Psychology* 50 (1): 169–74.
- Gardner, John. 2007. *Offences and Defences: Selected Essays in the Philosophy of Criminal Law*. Oxford University Press Uk.
- Handfield, Toby. 2008. "Unfinkable Dispositions." *Synthese* 160 (2): 297–308.
- Harman, Gilbert. 1999. "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian Society* 99 (1999): 315–31.
- Heil, John, and C. B. Martin. 1998. "Rules and Powers." *Philosophical Perspectives* 12 (S12): 283–312.
- Heim, Sarah. 2013. "The Applicability of the Duress Defense to the Killing of Innocent Persons by Civilians." *Cornell International Law Journal* 46 (1): 165–90.
- Herman, Barbara. 1981. "On the Value of Acting From the Motive of Duty." *Philosophical Review* 90 (3): 359–82.
- Hieronymi, Pamela. 2007. "Rational Capacity as a Condition on Blame." *Philosophical Books* 48 (2): 109–23.
- Hills, Alison. 2015. "The Intellectuals and the Virtues." *Ethics* 126 (1): 7–36.
- Hobart, R. E. 1934. "Free Will as Involving Determination and Inconceivable Without It." *Mind* 43 (169): 1–27.
- Hooker, Brad. 2000. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford University Press.
- . 2002. "The Collapse of Virtue Ethics." *Utilitas* 14 (1): 22.
- Hooker, Brad, and Margaret Olivia Little. 2000. *Moral Particularism*. Oxford University Press.

- Hume, David. 2000. *A Treatise of Human Nature*. Oxford University Press.
- Hurka, Thomas. 2006. "Virtuous Act, Virtuous Dispositions." *Analysis* 66 (289): 69–76.
- . 2013. "Aristotle on Virtue." In *Aristotelian Ethics in Contemporary Perspective*, edited by Julia Peters, 9. Routledge.
- Hursthouse, Rosalind. 1991. "Arational Actions." *Journal of Philosophy* 88 (2): 57–68.
- . 2013. "Virtue Ethics." Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/fall2013/entries/ethics-virtue/>.
- Husak, Douglas. 2005. "On the Supposed Priority of Justification to Excuse." *Law and Philosophy* 24 (6): 557–94.
- Hyman, John. 2014. "Desires, Dispositions and Deviant Causal Chains." *Philosophy* 89 (1): 83–112.
- Jacobson, Daniel. 2005. "Seeing by Feeling: Virtues, Skills, and Moral Perception." *Ethical Theory and Moral Practice* 8 (4): 387–409.
- Jaworska, Agnieszka. 2007. "Caring and Internality." *Philosophy and Phenomenological Research* 74 (3): 529–68.
- Johnston, Mark. 1992. "How to Speak of the Colors." *Philosophical Studies* 68 (3): 221–63.
- Kamtekar, Rachana. 2004. "Situationism and Virtue Ethics on the Content of Our Character." *Ethics* 114 (3): 458–91.
- Kelly, Erin. 2012. "What Is an Excuse?" In *Blame: Its Nature and Norms*, edited by Neal Tognazzini and D. Justin Coates. Oxford; New York: Oxford University Press.
- Kent, Bonnie. 2012. "Dispositions and Moral Fallibility: The UnAristotelian Aquinas." *History of Philosophy Quarterly* 29 (2).
- Korsgaard, Christine M. 1983. "Two Distinctions in Goodness." *Philosophical Review* 92 (2): 169–95.
- . 2008. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford University Press.
- Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language*. Harvard University Press.
- Langton, Rae. 2007. "Objective and Unconditioned Value." *Philosophical Review* 116 (2): 157–85.
- Levy, Neil. 2005. "The Good, the Bad, and the Blameworthy." *Journal of Ethics and Social Philosophy* 1 (2).

- Lewis, David. 1997. "Finkish Dispositions." *Philosophical Quarterly* 47 (187): 143–58.
- Manley, David, and Ryan Wasserman. 2007. "A Gradable Approach to Dispositions." *The Philosophical Quarterly* 57 (226): 68–75.
- . 2008. "On Linking Dispositions and Conditionals." *Mind* 117 (465): 59–84.
- Markovits, Julia. 2010. "Acting for the Right Reasons." *Philosophical Review* 119 (2): 201–42.
- Martin, C. B. 1994. "Dispositions and Conditionals." *Philosophical Quarterly* 44 (174): 1–8.
- McDowell, John. 1979. "Virtue and Reason." *The Monist* 62 (3): 331–50.
- McGrath, Sarah. 2011. "Skepticism About Moral Expertise as a Puzzle for Moral Realism." *Journal of Philosophy* 108 (3): 111–37.
- McKenna, Michael. 2012. *Conversation & Responsibility*. Oxford University Press.
- Mele, Alfred R. 2003. *Motivation and Agency*. Oxford University Press.
- Moore, G. E. 1903. *Principia Ethica*. Dover Publications.
- Moore, Michael S. 1997. *Placing Blame: A Theory of the Criminal Law*. Oxford University Press.
- Murdoch, Iris. 1970. *The Sovereignty of Good*. New York, Schocken Books.
- Murphy, Jeffrie G., and Jean Hampton. 1990. *Forgiveness and Mercy*. Cambridge University Press.
- Murphy, J. G. 1992. *Retribution Reconsidered: More Essays in the Philosophy of Law*. 1992 edition. Dordrecht Holland ; Boston: Springer.
- Nagel, T. 1976. "Moral Luck." *Aristotelian Society Supplementary Volume* 50 (226): 115–51.
- Nielsen, Kai. 2007. "There Is No Dilemma of Dirty Hands." In *Politics and Morality*, edited by Igor Primoratz, 1–7. Palgrave Macmillan.
- Pervin, Lawrence A. 1994. "A Critical Analysis of Current Trait Theory." *Psychological Inquiry* 5 (2): 103–13.
- Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy and Public Affairs* 13 (2): 134–71.
- Raz, Joseph. 1975. "Permissions and Supererogation." *American Philosophical Quarterly* 12 (2): 161–68.

- Richards, Norvin. 1987. "Acting Under Duress." *Philosophical Quarterly* 37 (146): 21–36.
- Rosen, Gideon. 2002. "Culpability and Ignorance." *Proceedings of the Aristotelian Society* 103 (1): 61–84.
- Ross, Lee, and Richard E. Nisbett. 1991. *The Person and the Situation: Perspectives of Social Psychology*. McGraw-Hill.
- Ross, W. D. 2002. *The Right and the Good*. Clarendon Press.
- Russell, Daniel C. 2009. *Practical Intelligence and the Virtues*. Oxford University Press.
- Ryle, Gilbert. 1949. *The Concept of Mind*. Hutchinson and Co.
- Scanlon, Thomas. 1998. *What We Owe to Each Other*. Belknap Press of Harvard University Press.
- . 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Belknap Press of Harvard University Press.
- Setiya, Kieran. 2007. *Reasons Without Rationalism*. Princeton University Press.
- . 2010. "Sympathy for the Devil." In *Desire, Practical Reason, and the Good*, edited by Sergio Tenenbaum, 82–110. Oxford University Press.
- Sher, George. 2002. "Blameworthy Action and Character." *Philosophy and Phenomenological Research* 64 (2): 381–92.
- . 2007. *In Praise of Blame*. Oxford University Press.
- . 2012. "Wrongdoing and Relationships: The Problem of the Stranger." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 49–65. Oxford University Press.
- Shoemaker, David. 2011. "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility." *Ethics* 121 (3): 602–32.
- . 2015a. *Responsibility From the Margins*. Oxford University Press.
- . 2015b. "Ecumenical Attributability." In *The Nature of Moral Responsibility*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith, 115–40. Oxford University Press.
- Shoemaker, David W. 2003. "Caring, Identification, and Agency." *Ethics* 114 (1): 88–118.
- Sidgwick, Henry. 1907. *The Methods of Ethics*. Thoemmes Press.
- Sliwa, Paulina. 2015. "Moral Worth and Moral Knowledge." *Philosophy and Phenomenological Research* 92 (2).

- Slote, Michael A. 2001. *Morals From Motives*. Oxford University Press.
- Smart, J. J. C., and Bernard Williams. 2013. *Utilitarianism: For and Against*. Cambridge University Press.
- Smith, Angela M. 2005. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115 (2): 236–71.
- . 2008a. "Character, Blameworthiness, and Blame: Comments on George Sher's in Praise of Blame." *Philosophical Studies* 137 (1): 31–39.
- . 2008b. "Control, Responsibility, and Moral Assessment." *Philosophical Studies* 138 (3): 367–92.
- . 2012. "Attributability, Answerability, and Accountability: In Defense of a Unified Account." *Ethics* 122 (3): 575–89.
- . 2015. "Responsibility as Answerability." *Inquiry* 58 (2): 99–126.
- Smith, Michael. 1994. *The Moral Problem*. Blackwell.
- Sreenivasan, Gopal. 2002. "Errors About Errors: Virtue Theory and Trait Attribution." *Mind* 111 (441): 47–68.
- Sripada, Chandra. 2016. "Self-Expression: A Deep Self Theory of Moral Responsibility." *Philosophical Studies* 173 (5): 1203–32.
- Stocker, Michael. 1976. "The Schizophrenia of Modern Ethical Theories." *Journal of Philosophy* 73 (14): 453–66.
- . 1989. *Plural and Conflicting Values*. Oxford University Press.
- Strabbing, Jada Twedt. 2016. "Attributability, Weakness of Will, and the Importance of Just Having the Capacity." *Philosophical Studies* 173 (2): 289–307.
- Stratton-Lake, Philip. 2000. *Kant, Duty, and Moral Worth*. Routledge.
- Strawson, Peter F. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48: 1–25.
- Sverdlik, Steven. 2011. *Motive and Rightness*. Oxford ; New York: Oxford University Press.
- Swanton, Christine. 2003. *Virtue Ethics: A Pluralistic View*. Clarendon Press.
- Tannenbaum, Julie. 2015. "Mere Moral Failure." *Canadian Journal of Philosophy* 45 (1): 58–84.

- Timpe, Kevin, and Craig Boyd. 2014. *Virtues and Their Vices*. Oxford University Press.
- Turns, D. 1998. "The International Criminal Tribunal for the Former Yugoslavia: The Erdemovic Case." *International and Comparative Law Quarterly* 47: 461–74.
- Vuoso, George. 1987. "Background, Responsibility, and Excuse." *The Yale Law Journal* 96 (7): 1661–86.
- Wallace, R. Jay. 1996. *Responsibility and the Moral Sentiments*. Harvard University Press.
- . 2006. *Normativity and the Will: Selected Papers on Moral Psychology and Practical Reason*. Oxford University Press.
- . 2011. "Dispassionate Opprobrium: On Blame and the Reactive Sentiments." In *Reasons and Recognition Essays on the Philosophy of T.M. Scanlon*, edited by Rahul Kumar, Samuel Freeman, and R. Jay Wallace. Oxford ; New York: Oxford University Press.
- Walzer, Michael. 1973. "Political Action: The Problem of Dirty Hands." *Philosophy and Public Affairs* 2 (2): 160–80.
- Watson, Gary. 1975. "Free Agency." *Journal of Philosophy* 72 (April): 205–20.
- . 1977. "Skepticism About Weakness of Will." *Philosophical Review* 86 (3): 316–39.
- . 1996. "Two Faces of Responsibility." *Philosophical Topics* 24 (2): 227–48.
- Williams, Bernard. 1978. "Politics and Moral Character." In *Public and Private Morality*, edited by Stuart Hampshire. Cambridge University Press.
- Williams, Bernard Arthur Owen. 1985. *Ethics and the Limits of Philosophy*. Harvard University Press.
- Wolf, Susan. 1980. "Asymmetrical Freedom." *Journal of Philosophy* 77 (March): 151–66.
- . 1990. *Freedom Within Reason*. Oxford University Press.
- . 2015a. "Character and Responsibility." *Journal of Philosophy* 112 (7): 356–72.
- . 2015b. "Responsibility, Moral and Otherwise." *Inquiry* 58 (2): 127–42.
- Zaibert, Leo. 2012. "On Forgiveness and the Deliberate Refusal to Punish: Reiterating the Differences." *Journal of Moral Philosophy* 9 (1): 103–13.
- Zimmerman, Michael J. 1987. "Luck and Moral Responsibility." *Ethics* 97 (2): 374–86.