Meaning, Decision, and Norms:

*Themes from the Work of Allan Gibbard*

Eds. Billy Dunaway and David Plunkett

Maize Books
1. Introduction

Most of the time we assume that we should be moral—we think that there are good practical reasons for us to do what morality requires us to do. But, especially when morality’s demands seem onerous, we can be tempted to waver in this belief, and crave a philosophical account of why there are such practical reasons to be moral. It is often assumed that this explanation will have to invoke substantive normative considerations, and many normative ethicists argue that their views provide the best explanation of why we should be moral. Contractarians like David Gauthier (1986), Kantians like Christine Korsgaard (1996), Contractualists like Tim Scanlon (1998), and Consequentialists like Peter Railton (1986) contend that if—but only if—their accounts of the demands of morality are correct, we can explain our reasons to be moral in terms of our allegedly more obvious practical reasons to promote our own self-interest, consistently value agency, act on principles no one can reasonably reject, or promote the well-being of others.

In this paper I argue to the contrary that the best explanation of why we should be moral is actually neutral about the content of morality. I contend that it is a conceptual truth that, if an act is required or recommended by morality, there are genuine practical reasons to perform it. I show how this follows from the best fitting attitude analyses of our moral concepts and a general relationship between fitting motives and practical reasons. While this account of the existence of practical reasons to be moral is neutral about what morality requires, it has important implications for normative ethics. By removing the explanation of why we
should be moral as a desideratum on normative ethical theories, it improves the prospects of
theories like Rossian Pluralism, which seem ill placed to give a unified morality-independent
explanation of why we should do what they say we are morally required to do.

My conceptual explanation of why there are practical reasons to do what morality requires
or recommends consists of two main parts: (1) an analysis of an act’s being required or recom-
mended by morality in terms of the appropriateness or fittingness of having a particular kind
of attitude towards that act, and (2) an account of the relationship between the fittingness of
motivational states and what there is reason to do. Following authors like A. C. Ewing (1939)
and Allan Gibbard (1990), I argue that analyzing moral concepts in terms of the fittingness
of moral emotions like guilt and anger provides the best explanation of what distinguishes
moral judgments from other normative judgments. But I contend that an act’s being required
or recommended by morality is best analyzed in terms of the fittingness of forward-looking
feelings of obligation to perform it, which involve motivation to perform the act.

Next, I argue that the existence of practical reasons to perform an act is a matter, not of
the act’s contributing to an end that one is motivated to pursue, but to an end that it is fit-
ting to be motivated to pursue. Since an act’s being required or recommended by morality
conceptually entails the fittingness of motivation to perform it, and the fittingness of this
motivation conceptually entails that there is reason to perform it, it is actually a conceptual
truth that there are reasons to do what is required or recommended by morality, whatever
that turns out to be.

Finally, I show how my account can explain why, although moral considerations are not
always overriding, we necessarily have conclusive reasons to do what morality requires. I
contend that an act counts as morally required only if the reasons to feel obligated to per-
form it are conclusive, which entails that it is unfitting to be most strongly motivated not to
perform it. This, together with my account of the connection between fitting motives and
practical reasons, entails that whatever considerations are weighty enough to make the act
morally required are conclusive reasons to perform it. This account goes deeper than that of
authors like Stephen Darwall (2006) and Douglas Portmore (2011), who claim that there are
conceptual connections between (1) an act’s status as wrong and its status as blameworthy,
and (2) its status as blameworthy and the existence of conclusive reasons not to perform
it. My account shows how an act’s status as morally wrong explains both the existence of
conclusive reasons to perform it and the connection between blame and practical reasons
to which such authors appeal.

While it is the business of substantive ethical theorizing to tell us which considerations if
any are weighty enough to make acts wrong, I argue that our reasons to believe that certain
acts (like inflicting massive suffering on innocents just for fun) are wrong are just as good as
our reasons to believe any other substantive normative claim (like that the fact that an act
would cause oneself pain is a reason to avoid it). What my metaethical account gives us is
an explanation of why our excellent reasons to believe that certain acts are wrong consti-
tute equally excellent reasons to believe that the considerations that make them wrong are
conclusive reasons not to perform them—without recourse to any further substantive ideas about the nature of practical reasons.

2. Moral Emotions and the Attractions of Fitting Attitude Analyses

Perhaps the greatest initial opposition to my approach will stem from the judgmentalist (or “cognitivist”) view that emotions like guilt and feelings of obligation themselves involve judgments about moral blameworthiness or wrongness, and thus cannot be used to informatively analyze moral judgments.¹ I thus begin by considering some problems with judgmentalism and some virtues of my alternative approach to the relationship between moral emotions and moral judgments.

A widely discussed problem for judgmentalism is the phenomenon of recalcitrant emotions, or emotions we feel but think unfitting or inappropriate. For instance, one can feel guilt in spite of the fact that one believes that one has done nothing blameworthy, and one can feel outrage or resentment towards someone despite the fact that one judges that she has done nothing wrong. Judgmentalism is committed to the view that having a recalcitrant emotion involves making a judgment in conflict with other judgments one holds. But merely having a recalcitrant emotion seems not to have to involve such a conflict in judgment. Suppose, for example, that I feel guilt for knocking over and breaking a friend’s lamp, though I exercised all due caution and think that I did nothing at all wrong or blameworthy. A conflict in judgment about whether I had done something culpable would involve such things as conflicting tendencies to draw inferences about the moral status of similar acts, conflicting views about whether I deserve reproach, and conflicting views about whether something is wrong with me for feeling what I feel.² But it seems that I can feel guilt about breaking the lamp and judge that the guilt makes no sense without any of these kinds of conflicts.

In response to this feature of recalcitrant emotions, those with judgmentalist leanings sometimes opt for a quasi-judgmentalist (or “perceptualist”) view according to which moral emotions like guilt involve “moral evaluations” that are something less than full-blown judgments, but still attribute moral properties like blameworthiness and wrongness.³ The

¹ For proponents of this view see, for example, Solomon (1976, 1988), Sabini and Silver (1982), and Foot (1959).
² Cf. D’Arms and Jacobson’s (2003, 129–30) discussion of the difference between merely fearing flying and judging flying dangerous, and the relevance of this claim to the judgmentalist’s need to posit inconsistent judgments wherever there are recalcitrant emotions. They discuss how those with phobic fears of flying “are typically well aware that [flying] is safer than activities they do not fear, such as driving to the airport . . . they do not worry when their friends fly, or buy insurance when forced to fly themselves,” concluding in their footnote 7 that “the great challenge for judgmentalist accounts of recalcitrant emotion is that the behavioral evidence supporting the attribution of the evidently suspect belief is problematic.”
³ For examples of quasi-judgmentalists treatments of this kind see, for instance, Roberts (1988) and Greenspan (1988). For criticisms of quasi-judgmentalism related to (as well as distinct from) those I present here, see Gibbard (1990, 39–40, 129–32) and D’Arms and Jacobson (2003).
quasi-judgmentalist idea seems to be that moral emotions involve states that are more akin to “moral perceptions” than moral judgments. In general, perceptual states might be distinguished from beliefs or judgments in that they are more “domain-specific” or less sensitive to learning and multiple sources of information, more quickly instanced, and incapable of being consciously inferred. But perceptual states also play an important role in inference processes by contributing their contents as “starting points” or data, which can of course be debunked by theories that best explain the totality of such contents, but in favor of which a burden of proof is set in inquiry. To the extent that we have “moral perceptions” that play these roles, I think that we may know them as “moral intuitions.”

The problem for quasi-judgmentalism, however, is that just as merely having a recalcitrant emotion seems not to have to involve conflicting moral judgments, it seems not even to have to involve a conflict between moral judgment and moral intuition. Return again to the guilt I feel for breaking my friend’s lamp, despite judging that I have done nothing blameworthy. It certainly seems that I can feel and judge this way without my having an intuition to the contrary—that is, without my having even spontaneous appearances to the effect that I deserve reproach or that there would be something wrong with me were I to fail to feel the guilt I do, and without any tendency to set a burden of proof in inquiry in favor of the view that my conduct was wrong.

But (quasi-) judgmentalists may face a problem even deeper than those posed by the phenomenon of recalcitrant emotions. This is that they must explain what these moral judgments are, which feelings like guilt, outrage, and resentment supposedly involve, without reference to these feelings themselves. In light of the wide diversity of things that people have coherently (though in many cases quite falsely) believed to be morally blameworthy or wrong, this seems to be a very difficult task. Some of these things include: inflicting harms upon others, failing to prevent harms to others, defecting in the presence of collective action problems, and failing to respect the autonomy of other agents. But they would also include all manner of apparently miscellaneous behavior, including sexual practices, drug use, violations of etiquette, “playing God” by engaging in cloning or genetic modification of organisms (quite apart from its effects on any individual’s well-being), failures to adhere to certain religious practices, stringing together certain phonemes (in the form of curse words), and so on. It should be emphasized that these kinds of apparent miscellany can and have been coherently thought to be intrinsically wrong or blameworthy quite apart from beliefs about their contribution to anyone’s well-being or autonomy.

Thus, an analysis of moral blameworthiness or wrongness in terms of an act’s failing to maximize happiness, or being hated by deities, or violating autonomy, or possessing any

---

4 For instance, Roberts (1988, 187–88) contends that such emotions involve “construing something in terms of a concept,” which he explains by reference to how ambiguous images, like the duck-rabbit, give rise to different perceptual states depending upon which concepts are tokened.

5 See, for example, Zimbardo and Weber (1997, chapter 5, especially 177–97).

6 Cf. Gibbard (1990, 130): “Anyone who claims that anger includes a judgment of moral transgression needs to explain the judgment.”
other substantive features would fail to account for how diverse coherent moral judgments can be, and what is at issue between people with rival moral views. Perhaps, however, we can explain what is common to all moral judgments by reversing the judgmentalist order of explanation and analyzing moral judgments as judgments of the fittingness of moral emotions. For instance, following such figures as John Stuart Mill and A. C. Ewing, Gibbard (1990, 40–45, 126–27) proposes the following analysis of the concept of moral blameworthiness:

**Fitting Attitude Analysis of Moral Blameworthiness:** To judge that what someone has done is morally blameworthy is to judge that it is fitting for her to feel guilty for having done it, and fitting for others to be angry at her for doing it.

All of the above coherent judgments that acts are blameworthy do seem to involve judging them to befit guilt and anger, and it seems difficult to identify anything else that they have in common.

One might wonder, however, what we gain by saying that common to all coherent moral judgments and disputes are views and disputes about which moral emotions are fitting, as opposed to saying that moral concepts simply resist being informatively understood in any further terms. What we seem to gain is an explanation of what moral judgments have in common with other normative judgments, like those concerning the fittingness, rationality, or appropriateness of desires, beliefs, and non-moral emotions. Common to all of these are views that a certain attitude is favored by reason, and the attitudes held to be favored by reason in the case of moral judgments are moral emotions like guilt and anger. These judgments about the fittingness of attitudes share such features as a wide diversity of things that can be coherently thought to befit them, our attempting to determine which of these coherent positions are correct via *a priori* methods of philosophical argument, and our conclusions about which responses are fitting exerting direct (non-behavior-mediated) causal pressure on our coming to have them. By subsuming these phenomena that we see in the case of moral judgments in relation to moral emotions under those of judgments of the fittingness of attitudes generally, we gain an (at least partial) explanation of them.

---

7 I should perhaps emphasize that by itself this in no way entails the falsity of any substantive view about which acts are morally wrong or blameworthy, including the utilitarian claim that the morally wrong acts are all and only those that fail to maximize happiness. The point here is simply that the utilitarian cannot intend her view as a conceptual analysis of wrongness; the denial of the position is coherent even if false, and we need an understanding of wrongness that can capture the substantive dispute between the utilitarian and her rivals.

8 On its being a distinguishing characteristic of judgments that attitudes are fitting (as opposed to supported by non-fittingness considerations) that they are capable of directly causing us to have them, see Gibbard (1990), Parfit (2001), Hieronymi (2005), and Raz (2009). As D’Arms and Jacobson (2009) argue, mere direct causal influence may be insufficient to distinguish fittingness from non-fittingness judgments in all cases. But as I suggest in the text (and argue in more detail elsewhere—see Nye 2009, chapter 6), judgments of fittingness are best understood as part of a psychological process characterized by several functional roles, which include but are not limited to directly influencing our attitudes.
3. Feelings of Obligation, Moral Wrongness, and Moral Reasons

In order to see how fitting attitude analyses can shed light not only on the nature of our moral concepts but also on the connection between morality and practical reasons, I believe that we must turn our attention from aretaic or hypological concepts like blameworthiness to deontic concepts of moral wrongness and moral reasons. After proposing to analyze judgments of moral blameworthiness as judgments about the fittingness of anger and guilt, Gibbard noted some ways in which blameworthiness and wrongness can come apart. For instance, if one lashes out in grief at a friend offering condolences, one’s conduct may be wrong but nevertheless exculpated by one’s overwhelming grief (Gibbard 1990, 44).

Gibbard concludes that “we need a distinct concept of wrong . . . as opposed to blameworthy,” noting that while the concept of blameworthiness is retrospective in character, the concept of wrongness is prospective. I think that the best way to understand this forward-looking character of the concept of moral wrongness is to see that it is concerned, not with the fittingness of guilt and anger towards what has already been done, but with the fittingness of the agent’s feeling obligated to do or avoid doing various things that it is open to her to do.

Feelings of obligation are, as Richard Brandt (1959, 117–18) observed, what you have when you see someone in trouble and feel like you “just can’t” leave her. J. S. Mill (1863) described the feeling as an “internal sanction of duty . . . a feeling in our own mind . . . attendant on violation of duty, which in properly cultivated moral natures rises, in the more serious cases, into shrinking from it as an impossibility,” and “a mass of feeling which must be broken through in order to do what violates our standard of right.” The phenomenology of feeling obligated not to do something is similar to that of feeling guilt for having done it, but whereas guilt is retrospective, feeling obligated not to do something involves a kind of prospective guilt-tinged aversion to doing it.9

As associated as these feelings of obligation may be with judgments that one is morally obligated to do something, it is possible to feel obligated recalcitrantly, or to feel obligated not to do things that one judges not to be wrong. For example, a man from a background with restrictive views about sexual morality might feel obligated not to engage in certain sexual practices even though he now thinks them perfectly morally permissible. Or a woman in an abusive relationship might feel obligated not to leave her partner, but be thoroughly convinced that she is in no way morally required to stay with him. As with our discussion of recalcitrant guilt above, it seems that the man and woman could in this way recalcitrantly feel obligated without any of the conflicting inferential tendencies, views about appropriate

---

9 It is important, however, to emphasize that feeling obligated not to do something involves an aversion to doing it, not to the prospect of feeling guilt for having done it. If you saw someone in need of help but had on hand a pill that would prevent you from feeling guilt for failing to help her, your feeling that you “just can’t” leave her (unlike an aversion to feeling guilt) would motivate you to help her and generate no motivation at all to take the pill.
conduct, or views about their own responses required for conflicting judgments about what they are morally obligated to do. It seems, moreover, that the man and woman could recalcitrantly feel obligated without any of the spontaneous appearances and tendencies to set burdens of proof in inquiry required for an intuition or sub-judgmental moral evaluation in conflict with their judgments about their moral obligations.

For reasons similar to those that favor analyzing judgments of moral blameworthiness as judgments about the fittingness of guilt and anger, I think that the content and normative force of judgments that acts are wrong or opposed by moral reasons are best captured by analyzing them as judgments about the fittingness of feeling obligated not to perform them. For instance, what seems distinctive about viewing the fact that doing $A$ will save someone’s life as a moral reason to do $A$ is one’s taking this consideration to count in favor of feeling obligated to do $A$. Similarly, what seems distinctive about thinking that the fact that doing $A$ would kill someone makes it morally wrong or forbidden (as opposed to just unreasonable) to do $A$ seems to be one’s taking this consideration to make it, on balance, fitting for you to feel obligated not to do $A$.

This supports the following analyses of our concepts of moral reasons and moral wrongness:

**Fitting Attitude Analysis of Moral Reasons**: To judge that $R$ is a moral reason for agent $X$ to $\phi$ is to judge that $R$ is a fittingness reason for $X$ to feel obligated to $\phi$, and

**Fitting Attitude Analysis of Moral Wrongness**: To judge that it is morally wrong for $X$ to $\psi$ is to judge that it is, on balance, fitting for $X$ to feel obligated not to $\psi$.

---

10 The best alternative proposal about what is distinctive about viewing this as a moral reason is something like that it involves one's taking it to be a reason that one has simply because one's act will promote the well-being of the individual in question. But it is surely coherent to think that there are distinctly moral reasons to do things other than promote well-being: with some plausibility one can think there are intrinsic moral reasons to respect autonomy and keep promises, and we know only too well what someone is thinking when she takes the alleged fact that an act is “unnatural,” “against tradition,” or “against God's will,” to be an intrinsic moral reason against doing it. Moreover, although most of us are decent enough to accept a substantive principle of beneficence according to which there is intrinsic moral reason to promote the well-being of every individual capable of well-being, it is, sadly, coherent to think otherwise. The view that there are individuals whose well-being there is no intrinsic moral reason to promote (although perhaps still some intrinsic non-moral reason to promote) has been coherently entertained, for instance, by some who take exalted views of the moral relevance of such factors as retribution, autonomy, promise-keeping, supernatural wills, and group-loyalty.

11 Of course, we can think it perfectly fitting for someone to experience no feelings of obligation to refrain from doing things we think wrong if she is already sufficiently motivated not to do them. In most cases we would never even consider doing things that would kill others, and if we do, care for those others and fear of punishment are almost always sufficient deterrents. Although we think it would be wrong for us to kill in such cases we surely do not think it inappropriate that we experience no feelings of obligation to refrain from doing so. Moreover, as I discuss in more detail below in section 5, there is a sense in which we can
Just as judgments about the blameworthiness of actions have the central normative feature of guiding feelings of guilt and anger towards them, judgments about wrongness and moral reasons seem to have the central normative property of guiding feelings of obligation. These fitting attitude analyses of moral judgments can explain their ability to generate motivation to act out of feelings of obligation as a special case of the ability of judgments that attitudes are fitting to directly guide us into having them.12

Finally, these analyses can help explain the gap Gibbard noted between judging an act wrong and judging it blameworthy. Combining them with Gibbard’s existing analysis of blameworthiness, we can understand, say, thinking that someone’s lashing out in grief was wrong but not blameworthy as a thought to the effect that although it isn’t fitting for us to feel angry at the person who lashed out and it isn’t fitting for her to feel guilt for lashing out, it still was the case that before she lashed out she should have felt obligated not to do it.13

Think it fitting on balance to feel obligated to do things that we do not think it wrong to fail to do. It seems perfectly fitting for someone who goes above and beyond what morality requires, say by getting killed to save a younger stranger from death, to feel obligated to do what this.

To clarify my proposal, it is important to note first that talk of feeling emotions, like talk of desiring or preferring, is ambiguous between an occurrent and a dispositional sense. Occurrent feelings and preferences exert causal pressure on one’s behavior at the moment, and (at least typically) involve phenomenal experiences, while dispositional feelings and preferences merely have the disposition to become current in certain circumstances. Thus, one can dispositionally feel obligated not to push one’s friends out of windows in the same way one can dispositionally feel anger at one’s father even while one is enjoying his company and experiencing no negative emotions. Second, as I discuss below in section 5, it is important to note that a response’s being “fitting on balance” is ambiguous between (1) the response’s being mandatory, in that there is no alternative response that is as strongly supported by fittingness reasons, or (2) the response’s being justified, in that there is no alternative response that is more strongly supported by fittingness reasons.

In more detail, then, my proposal is that to think it morally wrong for $X$ to $\psi$ is to think that it is mandatory for $X$ to have at least a dispositional feeling of obligation not to $\psi$ (and mandatory for $X$ to have an occurrent feeling of obligation not to $\psi$ only if $X$ is not already sufficiently motivated not to $\psi$). The sense in which one can judge it “fitting on balance” for $X$ to feel obligated to $\varphi$ when one takes $X$’s $\varphi$-ing to be supererogatory is that one thinks $X$’s feeling of obligation is justified but not mandatory.

To appreciate the centrality of this attitude-guiding role of moral judgments, suppose that someone were to label as “morally wrong” all those things we would call wrong, but took this to have no significance for what it was appropriate to feel obligated to do and consequently had no propensity to feel obligated not to perform the acts in question. It seems that by “wrong” she would not really mean wrong. On the other hand, if someone were to label as “morally wrong” precisely those things we think permissible, she would still seem perfectly intelligible as thinking that those things are wrong so long as she thought it was fitting to feel obligated not to perform them.

One might be wondering, however, why there is not in addition to a coherent wrongness-blameworthiness gap a similar coherent blameworthiness-wrongness gap. As we have seen, it seems perfectly coherent to think that it is fitting to feel obligated not to perform an act but that it is also unfitting to feel guilt or for others to feel outrage at one for performing it. But it seems incoherent to think that it is fitting to feel guilt or for others to feel outrage at one for performing an act if it was not fitting for one to feel obligated not to perform it in the first place. One thing I should point out is that conceptual connections between the fittingness of different moral emotions are already an issue for fitting attitude analyses like Gibbard’s analysis of moral blameworthiness, in that their proponents need to explain why it seems incoherent to hold that
4. Fitting Attitudes and Reasons to Act

I have thus argued that to judge an act wrong or opposed by moral reasons is to judge that there are considerations that make it fitting to feel obligated not to perform it. Since a judgment’s truth entails the truth of its analysans, this means that it is a conceptual truth that (1) \( R \) is a moral reason for \( X \) to \( \varphi \) iff \( R \) is a fittingness reason for \( X \) to feel obligated to \( \varphi \), and (2) \( X \)'s \( \psi \)-ing is morally wrong iff it is fitting for \( X \) to feel obligated not to \( \psi \).\(^{14}\) I will now argue that these analyses, together with general facts about the relationship between fitting attitudes and reasons to act, explain why an act’s deontic status entails the existence of practical reasons to perform or avoid performing it.

The basic idea here is that what there is reason for us to do is determined by what aims there is reason for us to have, and the question of what aims there is reason to have is identical to that of what it is fitting to be motivated to do. Since attitudes like feeling obligated to \( \varphi \) (or to avoid \( \psi \)-ing) involve motivation to \( \varphi \) (or avoid \( \psi \)-ing), the fittingness of these attitudes acts can befit outrage on the part of others without befitting guilt on the part of their performers, or that acts can befit guilt on the part of their performers without befitting outrage on the part of others. I think that the answer in both cases is that it is a conceptual truth about these attitudes that states would not count as guilt, outrage, or feelings of obligation unless their fittingness was interrelated in these ways.

Although the details are beyond the scope of this paper, here very briefly is how I think we came to have emotion concepts like this. For evolutionary reasons, our ancestors tended to feel guilt and outrage only towards acts that were such that their performers would tend to feel obligated not to perform them the first place—but to tend not to feel outrage or guilt towards all such acts the performers of which would tend to feel obligated not to perform (when the performance was, e.g., due to overpowering impulses). When the governance of emotions by norms came on the scene, similar evolutionary pressures favored our ancestors’ accepting systems of norms that prescribed feeling guilt and outrage only towards acts they also prescribed feeling obligated not to perform in the first place (but not vice versa). This was such a central feature of our ancestors’ systems of norms for emotions that the folk psychological theory that came to be true of us was thus one according to which the states that played the guilt, outrage, and feeling of obligation roles were such that the first two were prescribed by norms only when the last was (but not vice versa, and also such that each of the first two were prescribed by norms only when the other was). Because such a folk theory was true of us (and we weren’t too dim), it was the folk theory we came to have, and from which our folk emotion concepts of guilt, outrage, and feelings of obligation were extracted via the Ramsey-Carnap-Lewis method of analyzing theoretical concepts (see, for instance, Lewis (1970, 1972). For more on this suggestion about the origin of our emotion concepts as it relates to conceptual connections between the fittingness of different emotions, see Nye (2009, chapter 5).

\(^{14}\) Compare: if judging someone to be a bachelor amounts to judging him to be a male who is not in a romantic relationship but in a position to enter one, then it is a conceptual truth that someone is a bachelor iff he is a male who is not in a romantic relationship but in a position to enter one. Because analyses of one kind of judgment into another in this way support analytic relationships between the facts the judgments represent, I will slide rather freely between talking about what it is to make a certain kind of judgment (“to judge an act wrong is to judge that it is fitting to feel obligated not to perform it”) and talking about the analytic relationships between the facts they represent (“it is a conceptual truth that an act is wrong iff it is fitting to feel obligated not to perform it”).
entails the fittingness of this motivation, which entails the existence of reasons to \( \varphi \) (or to avoid \( \psi \)-ing).

The first part of this connection between fitting attitudes and reasons to act can be stated as a

**Warrant Composition Principle** [WCP]: Let \( M \) be a mental state that involves mental state \( M' \) as an essential component. If \( R \) is a fittingness reason to be in \( M \), then \( R \) is a fittingness reason to be in \( M' \).

WCP simply states that if there is reason to be in a mental state, then necessarily there is reason to be in all that the state essentially involves. For instance, if one acknowledges my claim that part of what it is to feel obligated to \( \varphi \) is to be motivated to \( \varphi \), it would seem incoherent to hold that a consideration (like \( \varphi \)'s relieving someone's pain) counts in favor of feeling obligated to \( \varphi \) but does not count in favor of being motivated to \( \varphi \). Since having the motivation is simply part of what it is to have the feeling of obligation, a consideration cannot make the feeling of obligation fitting without making the motivation fitting as well.

The second part of this connection between fitting attitudes and reasons to act is the relationship between what it is fitting to be motivated to do and what there is reason to do, which can be stated as a

**Motivations-Actions Principle** [MAP]: Let \( \varphi \)-ing be an action. If \( R \) is a fittingness reason to be motivated to \( \varphi \), then \( R \) is a reason to actually \( \varphi \).

Just as the consideration that \( \varphi \)-ing would relieve someone's suffering cannot make it fitting to feel obligated to \( \varphi \) without making it fitting to be moved to \( \varphi \), so too it seems this consideration cannot make it fitting to be moved to \( \varphi \) without actually counting in favor of \( \varphi \)-ing.\(^{15}\)

It is intuitive that what there is reason to do is determined by what aims there is reason to have. I think that the best theoretical explanation of MAP is that, because practical reasoning governs our actions by means of governing our motives, the process of determining what aims to have—and thus what to do—is essentially a process of determining what intrinsic motives to have. As authors like Michael Bratman (1987, 54) and Thomas Scanlon (1998, 20–21) have argued, because our practical reasoning controls our actions by controlling our intentions to perform them, reasons to perform an action just are reasons to intend to perform it. But it must be clarified that reasons to do \( A \) are identical to fittingness reasons to intend to do \( A \). As Kavka's (1983) toxin puzzle illustrates, merely pragmatic reasons to intend to do something (like the reason to intend—or get oneself to intend—to drink a toxin

\(^{15}\) WCP and MAP closely resemble John Skorupski's principles FDF and FDD, the conjunction of which he referred to as the "Feeling / Disposition Principle" (1999, 38, 63, 131, 174 n24) and more recently as the "Bridge Principle" (2010, 265–67).
tomorrow constituted by the fact that a reliable mind reader will pay you now if you intend
this) need not be reasons to actually do it.

Moreover, in light of the role intentions play in realizing the objects of our desires and
other valenced attitudes (like feelings of obligation), there are similar reasons to think that,
because reason ultimately governs our intentions by governing these attitudes, the fitting-
ness of intentions is itself determined by the fittingness of these other motives. As Bratman
argues, the role of intentions is not to supply an utterly new source of motivation that con-
flicts with the motives involved in our valenced attitudes, but to help cognitively limited
agents like us realize the objects of these motives over time. This role of intentions entails that
their normative assessment must be tied closely to that of the valenced attitudes they serve.

Although Bratman often speaks as though practical reasoning must simply take our
intrinsic valenced attitudes as given, it seems clear that we can assess them as reasonable
or unreasonable by determining through philosophical reasoning whether they are fitting or
unfitting. Moreover, as we have seen, it is characteristic of these fittingness assessments
that they directly guide our attitudes. For instance, one might start out with much weaker
feelings of obligation to avoid inflicting harms of a given size (like a given amount of suf-
fering), independent of their further consequences, upon non-human animals than upon
humans. But one might then reflect upon how bare biological species membership amounts
merely to something like a shared history of phylogenetic descent, phenotype-independent
genotype, or psychology-independent morphology, and how, with reference to profoundly
intellectually disabled humans, someone's lesser intellectual ability does not seem to justify
lesser concern with her equally-sized interests. As a result, one might come to judge one's
initially weaker feelings of obligation to avoid inflicting given-sized harms on non-human
animals to be unfitting. This judgment that one's feelings are unfitting tends directly to
change them, and to alter one's intentions from carrying out one's previous aim of, say,
avoiding the infliction of given-sized harms on non-human animals only when one finds this
relatively convenient (consistent, for instance, with one's previous intentions to continue to
consume animal products) to carrying out one's new aim of avoiding inflicting such harms
on non-human animals with as much priority as one gives to the avoidance of inflicting them
upon humans (issuing for instance, in a new general intention to consume a vegan diet, and
a new particular intention to buy vegan food at the grocery store).

Thus, because reason governs motives other than intentions through determinations of
their fittingness, and intentions are simply a means of achieving the objects of these motives,
fittingness reasons for these other motives are identical to fittingness reasons for intention.
The role of intentions is primarily to enable us to settle in advance what future courses of
action will best achieve the ends that it is fitting to be most motivated to achieve (like being

---

16 While Bratman often speaks as though our intrinsic desires or pro-attitudes are themselves normative reasons for intention and action, he makes it clear that he actually wishes to remain neutral between this view and the view that our intrinsic desires can be assessed as reasonable or unreasonable (1987, 22).
vegan to minimize harm to non-human animals), and to pick from among the many courses of action that often have equally good prospects of doing this (like going to one of two otherwise equally choiceworthy grocery stores to purchase vegan food). Together with the above observation that reason governs our actions through determinations of the fittingness of the intentions that lead us to perform them, this entails that, because reason ultimately governs our intentions and actions by determining the fittingness of the motives they seek to serve, fittingness reasons to be motivated to do something are identical to fittingness reasons to intend to do it and practical reasons to do it.\footnote{It is important to clarify that neither MAP nor this explanation of it makes what there is reason for us to do dependent upon what attitudes or motives we actually happen to have. The idea is that what there is reason for an agent to do is what would serve the objects of fitting intrinsic motives, by which I mean the intrinsic motives that it \textit{would be fitting} for her to have, \textit{whether she has them or not}. On this view, if a consideration (like a policy’s preventing suffering) is a fittingness reason for an agent to be motivated to do something (like vote for the policy if she can), then it is a reason for the agent to do this regardless of whether she ever has or comes to have any actual motivation to do it.}

Having thus argued in favor of my fitting-attitude analyses of moral wrongness and reasons, WCP, and MAP, I can use them to give the following explanation of why, if a consideration is a moral reason to do something, then it is also a practical reason to do it, and why, if an act is morally wrong, then one has practical reason not to perform it. Since feeling obligated to \( \varphi \) essentially involves motivation to \( \varphi \) (and feeling obligated not to \( \psi \) essentially involves motivation not to \( \psi \)), it follows from (1) and (2) together with the WCP that (1’) if \( R \) is a moral reason to \( \varphi \), then \( R \) is a fittingness reason to be motivated to \( \varphi \), and (2’) if \( \psi \)-ing is morally wrong, then it is fitting to be motivated not to \( \psi \). Moreover, it follows from (1’) and (2’) together with the MAP that (1*) if \( R \) is a moral reason to \( \varphi \), then \( R \) is a genuine practical reason to \( \varphi \), and (2*) if \( \psi \)-ing is morally wrong, then there are genuine practical reasons not to \( \psi \).

5. Conclusive Reasons Not to Do Moral Wrong

If my strategy for explaining moral concepts and their relation to reasons for action is correct, we are thus guaranteed a conceptual connection between an act’s being favored by moral reasons and our having practical reason to perform it. Our having moral reason to do something seems to entail that we have some practical reason to do it, but not conclusive
reason. Some acts, like getting oneself shot in the head to save a slightly younger stranger from drowning, seem to be favored by moral reasons, but seem not to be morally required, or are not morally wrong to fail to perform. In the case of such an act, the moral reasons to perform it (viz., that it would save the stranger’s life) seem to constitute practical reasons to perform it. But it at least seems coherent to think that one lacks conclusive practical reason to perform such an act, or that one has sufficient practical reasons not to perform it (constituted, e.g., by the fact that the act will get you killed).

But the notion of moral wrongness or moral requirement (where what is morally required is just what it is morally wrong not to do) seems to be different in this respect. There seem to be genuine problems with the coherence of thinking that one's doing something would be morally wrong but that one has sufficient reason to do it anyway. As such, a strong version of what Stephen Darwall (1997, 306) calls the thesis of “morality-reasons internalism” might seem to be true of moral requirement, namely “if $S$ is morally required to do $A$, then necessarily there is conclusive reason for $S$ to do $A$.”

It would be odd, however, if this strong thesis were to be explained solely in terms of the weightiness of the considerations that make acts morally required or wrong not to perform. Such considerations—for instance that I have promised to be across town and that she will die if I don't stay and help—can be brought into conflict without necessarily giving rise to rational dilemmas,¹⁸ and it seems at least coherent to think that they are at times outweighed by non-moral reasons like getting across town will get me killed.¹⁹ As such, I think that a much more attractive explanation of the strong thesis is that whether an act gets to count as falling under our concept of moral requirement—unlike, say, our concept of simply being favored by moral reasons—is itself sensitive to whether or not the reasons in favor of performing it are actually conclusive. That is, as W. D. Falk (1948, 124) suggested, “our very thinking that we ought [that is, are morally required] to do some act already entails that, by comparison, we have a stronger reason in the circumstances for doing it than any other.”

I think that the strategy I have been pursuing for analyzing moral concepts and explaining their connections to reasons for action can help explain why Falk’s kind of account of the necessary conclusivity of reasons not to do moral wrong is in fact correct. First, I should clarify the kind of fittingness reasons for feeling obligated not to do things with which the fitting attitude analysis identifies their moral wrongness. The idea is not that an act is morally wrong if one is simply justified or rationally permitted to feel obligated not to perform it in the same way in which one is justified or permitted to feel angry at actors whose conduct is

---

¹⁸ By which I mean situations in which whatever one does is irrational or other than one has most reason to do. I, for the record, do not think that it is conceptually possible for there to be such situations.

¹⁹ If the reader thinks that duties to oneself render this a moral reason, I invite her to consider whether there is some degree of trivialness of promise and some degree of harm that will befall one if one keeps it such that it is at least coherent to think that: (1) were it not for the harm to oneself one would be morally required to keep the promise, but (2) given the harm one would incur by keeping the promise it is rationally permissible to break it, yet (3) one does not “owe it to oneself” to prevent the harm to oneself by breaking the promise.
blameworthy. That an act is blameworthy does not entail that others are necessarily feeling something unfitting if they fail to feel angry with the blameworthy actor. Especially if the transgression is slight, it might be perfectly fitting for others to fail to feel such anger if they have more important matters to tend to, if the blameworthy actor is very remote, or if the blameworthy actor has done her best to make amends for what she has done or enough time has passed. 20 The kind of reasons one has to feel obligated not to perform morally wrongful acts, however, are not so easily overridden. Rather, to think an act morally wrong seems to involve thinking that, unless one is already going to refrain from performing it (for instance, one is sufficiently motivated not to perform it, or simply not motivated to perform it), one has conclusive fittingness reason to feel obligated not to perform it, in the sense that it would be unfitting for her not to feel so obligated. 21

Next, I would argue that while conclusive reasons to be in motivational states like emotions need not always amount to conclusive reasons to act out of them, conclusive reasons to feel obligated to do things are atypical in this respect. Consider, for instance, our reasons for wanting and against eating good tasting but unhealthy foods. That the foods taste so good seems to make it fitting to want to eat them. That the foods are unhealthy seems to make it fitting to be averse to eating them, and to be a practical reason to refrain from eating them. But the unhealthiness of the foods does not seem to make it unfitting to want to eat them at all. In some cases, we take the latter set of reasons to be weightier, and think that, all things considered, we should not eat the foods. But since these reasons can leave intact our reasons to want to eat the foods, it seems that in such cases we can have that (1) it is fitting to be somewhat motivated to eat the foods, (2) it is fitting to be more strongly motivated not to eat them, and (3) there is most practical reason to act out of our motives not to eat the foods and refrain from doing so.

What this kind of case suggests is that we need to distinguish between having conclusive fittingness reason to have some motivation to perform an act and having conclusive fittingness reason to be more strongly motivated to perform an act than any of its alternatives. Let us call the former states of having some motivation “gradational motivations” to perform the act, and call the latter state of strongest motivation a state of being “most motivated” to perform it. States of being most motivated to perform an act are those that arise as a result of the combined strengths of one’s gradational motivations to perform it being greater than the combined strengths of one’s gradational motivations not to perform it. The connection between conclusive reasons for motivation and action suggested by the above example, then,

---

21 Similarly, I should think that the reasons a morally blameworthy actor has to feel retrospective guilt for what she has done are not so easily overridden as those of others to feel angry at her. To think an act blameworthy seems to involve thinking that, at least until its performer has made amends or enough time has passed, and unless significantly more pressing matters arise (which must be more pressing than those minimally necessary to permit others not to feel angry at her), its performer has conclusive reason to feel guilt for performing it.
is that while one can have conclusive fittingness reason to be gradationally motivated to do something without having conclusive reason to do it, one’s having conclusive fittingness reason to be most motivated to do it entails that one has conclusive reason to do it. Call this the most-motivation-action principle:

**Most-Motivation-Action Principle:** If \( S \) has conclusive fittingness reason to be most motivated to do \( A \), then \( S \) has conclusive practical reason to do \( A \).

We have thus seen how for some gradational motivations, fittingness reasons to be in conflicting motivational states do not themselves constitute fittingness reasons against having these gradational motivations at all. It is in this respect, however, that feelings of obligation to do something seem to be different. Fittingness reasons to be motivated not to do something actually do seem to count against the fittingness of feeling obligated to do it. For instance, consider a situation in which one must break a promise in order to save someone’s life. The fact that one has promised to do something is a reason to feel obligated to do it. In this case, however, we would seem to take the fact that one must not do what one has promised to do to be a fittingness reason, not only in favor of being most motivated not to keep one’s promise, but indeed against feeling obligated to keep it under the circumstances.22

---

22 It is important here to distinguish the fittingness reasons to feel obligated to do things of which I am speaking from some closely related phenomena. To borrow (and use for slightly different purposes) an example from D’Arms and Jacobson (1994, 742–43), one’s mother might deeply fear being put in a nursing home, though given one’s inability to care for her and the costs to other family members one has most reason to put her in a home. In such a case, it might seem consistent with thinking that one has conclusive reason to put mother in the home to think that there is something wrong with one if one does not feel obligated to omit putting her in the home. Similarly, it might seem consistent with one’s having conclusive reason to put her in the home that it would be unfitting for one to feel no kind of reluctance towards putting her in the home, or to be able to put her in the home “with perfect equanimity” (my thanks to Stephen Darwall for this way of putting the intuition).

I contend, however, that these are not thoughts about the consistency of thinking that one has conclusive reason to put one’s mother in the home with thinking that one has conclusive fittingness reason to feel obligated not to do so. The first is most likely a thought that it is morally bad or disestimable to fail to feel obligated to omit putting her in the home. Such a judgment may easily be mistaken for a judgment that one has conclusive fittingness reason to feel so obligated because the former resembles the latter in two important respects: (1) it entails that there is reason to be motivated to do what one would do if one felt the aversion, and (2) for reasons discussed by Velleman (2002), it can translate into feeling obligated to omit putting mother in the home without one’s having to do anything to bring this about (though in a way that is dependent on judging the disesteem fitting).

The second thought is most likely a thought that one has conclusive fittingness reason to feel an attitude that we might call compunction, which bears some similarities to but can still be distinguished from the feelings of obligation the conclusive fittingness of which I am claiming we can understand moral wrongness in terms of. Phenomenally, compunction might also seem in a sense to be “guilt tinged,” but feeling compunction towards performing an act seems to involve something more like a feeling of hesitancy about, being unsettled about, or reluctance about performing it. What I have been calling feelings of obligation not to perform an act do not seem so aptly characterized in these ways—they seem to involve something
This apparent fact that any considerations that count in favor of being most motivated not to do something must also count against feeling obligated to do it would entail that one's reasons to feel obligated to do something can only be conclusive if they outweigh the reasons in favor of being most motivated not to do it. This in turn entails the following thesis:

**Contour Thesis:** If $S$ has conclusive fittingness reason to feel obligated to do $A$, then

$S$ has conclusive fittingness reason to be most motivated to do $A$.\(^{23}\)

We can now combine these considerations in favor of the most-motivation-action principle and the contour thesis with the fitting attitude analysis of moral wrongness to vindicate the strong morality-reasons internalism thesis about moral requirement. Given our clarification of the fitting attitude analysis, if it would be morally wrong for $S$ to fail to do $A$, then either $S$ is already sufficiently moved to do $A$ or $S$ has conclusive fittingness reason to feel obligated to do $A$. If $S$ is already sufficiently moved to do $A$, then either $S$ is in a fitting state of being most motivated to do $A$, or $S$'s state of being most motivated to do $A$ is unfitting. In the latter case (in which, for instance, $S$ is only motivated not to do $A$ because she irrationally believes that she will be punished for doing $A$), reason requires that $S$ cease to be in this state and thus be such that she has conclusive fittingness reason to feel obligated to do $A$. By the contour thesis, if $S$ has conclusive fittingness reason to feel obligated to do $A$, then $S$ has conclusive fittingness reason to be most motivated to do $A$. Thus, if it would be morally wrong for $S$ to

---

\(^{23}\) One might wonder why the contour thesis is, as I have argued it seems to be, true, or what makes (or guarantees that something makes) it true. I think that it is a conceptual truth about feelings of obligation that reasons to act contrary to such feelings count against having them at all. Were an attitude to be otherwise similar to these feelings in terms of phenomenology, attention direction, and motivation, but were to lack this feature, I think it would still fail to count as our feeling of obligation, conclusive reasons for which constitute the moral requirement to perform (or the moral wrongness of failing to perform) its object.

Although the details are again beyond the scope of this paper, I think that we came to have an emotion concept like this for much the same reasons as those I mentioned in footnote 13. That is, for evolutionary reasons, our ancestors came to tend to feel this kind of guilt-tinged aversion as an adaptive inhibition to defecting, where it was important to our genes (as it were) that we were conclusively so deterred. See, for example, Kitcher (1998, esp. 299–303) for a discussion of how fragile cooperation can be when inhibitions against defection are just one motive in the “internal melee” among many, as with our evolutionary cousins, the chimpanzees. When the governance of emotions by norms came on the scene, there were similar evolutionary pressures for our ancestors to come to accept norms that required feeling obligated to do something only when they required no stronger (or equally strong) motives to the contrary. This was such a central feature of their systems of norms that the folk psychological theory that came to be true of us was one where the states that played the feeling of obligation role were ones that were prescribed by norms only when no stronger (or equally strong) motives to the contrary were prescribed. Because it was true of us and we picked up on it, this was the folk psychological theory we came to have and over which we Ramsified to arrive at our folk concept of feelings of obligation. For more on this see Nye (2009, chapter 5).
fail to do \( A \), reason will only allow \( S \) to be most motivated to do \( A \), so \( S \) has conclusive fittingness reason to be so motivated. Finally, by the most-motivation-action principle, if \( S \) has conclusive fittingness reason to be most motivated to do \( A \), then \( S \) has conclusive reason to do \( A \). Thus, if it would be morally wrong for \( S \) to fail to do \( A \), which is to say that \( S \) is morally required to do \( A \), then \( S \) has conclusive practical reason to do \( A \).24

6. Conclusion

So if one thinks that one is morally required to do \( A \) and still wonders, “why should I do \( A \)?,” the answer is simply: “whatever makes it the case that you are morally required to do \( A \).” What my account gives us is an explanation of why the fact that these considerations are moral-requirement-makers guarantees that they are also conclusive practical reasons to act as one is morally required to act.

It may be important to conclude by emphasizing what my kind of vindication of reasons to be moral does and does not show. On my account, we have conclusive practical reason to do what we are morally required to do only because: (1) for \( S \) to be morally required to do \( A \), it must be the case that, absent a sufficient motivation or tendency to do \( A \) anyway, \( S \) has conclusive fittingness reason to feel obligated to do \( A \), and (2) in contrast to reasons for some other motivational states, if the fittingness reasons that count in favor of \( S \)’s feeling obligated to do \( A \) are not sufficiently weighty to determine that she has conclusive reason to act out of them, they are also insufficiently weighty to determine that she has conclusive fittingness reason to have the feeling of obligation. Like other kinds of reasons, an agent’s reasons to

24 I should point out that if one doubts the truth of the contour thesis for reasons related to cases like that discussed in footnote 22, there are still at least two other ways in which my approach to moral concepts and their connection to reasons for action can help to vindicate Falk’s kind of account of why we have conclusive reason not to do what is morally wrong. One might wish to insist that it is consistent with thinking that one has conclusive reason to put mother in the home that one has conclusive reason to feel (not just compunction but) some feeling of obligation not to do so. It could be argued, however, that this is only consistent with the thought that one has conclusive reason to put mother in the home because one also thinks that one has reason to feel an even stronger feeling of obligation to put her in the home (my thanks to Stephen Darwall for making me aware of this option). It could moreover be argued that the thought that one should feel most strongly obligated to put mother in the home entails both that one is morally required to do so and (due to the sensitivity of the fittingness of strongest feelings of obligation to countervailing considerations) that one has conclusive reason to be most motivated to put her in the home.

If, however, one is reluctant to accept my defense of the original contour thesis in relation to cases like putting mother in the home, one might be apt to object to the above modification in slightly altered cases for similar reasons. For instance, suppose that one has made a reasonably important promise to a friend that turns out to be extremely personally costly to keep. One might think that it is consistent with thinking that in such a case one has most reason not to keep the promise that one still has conclusive reason to feel (not just compunction but) some feeling of obligation to keep it, and that this is the only thing it makes sense to feel obligated to do in the circumstances. I would still at least contend that the thought that one’s feelings of obligation to keep the promise are rationally overpowered by motivations to break it entails that it would not be morally wrong to break the promise.
do things out of feelings of obligation can be overwhelmed by other considerations; it is simply that when they are, they are overwhelmed on the front of determining what it is fitting for the agent to feel and hence (given the fitting attitude analysis of moral wrongness) are no longer sufficient to make it morally wrong for the agent to do otherwise, or morally required for her to do what they are reasons to do.

My vindication of the strong morality-reasons internalism thesis thus does not give any conceptual guarantee that any given consideration—even *I have promised* or *she will die if I don't help her*—is either a genuine moral-requirement-making feature or a weightier reason than any other. It would show, however, that if we are as a matter of substantive normative fact morally required to do something, then we have conclusive practical reason to do it. I think that our reasons to believe that we are morally required to do things are just as good as our reasons to believe that other normative concepts are instantiated, and that this consists in the best unification and explanation of what substantive normative principles are most plausible after ideal reflection on what they are really saying. To the extent that our best such normative theories tell us that we are morally required to do something, I think we should conclude that we are, which I have argued entails that we have conclusive practical reason to do it.

What my conceptual account gives us is an explanation of why our rock-solid reasons to believe that certain acts are wrong constitutes equally rock-solid reasons to believe that the considerations that make them wrong are conclusive practical reasons not to perform them. As such, we do not need an independent substantive account of why we should care about what morality requires us to do. We do not have to rely upon substantive morality-independent intuitions to the effect that we have practical reasons to promote our own self-interest, consistently value agency, act on principles no one can reasonably reject, or promote the well-being of others to justify our practical reasons to be moral if but only if morality requires us to do what Contractarians, Kantians, Contractualists, or Consequentialists say it does. Quite apart from any such substantive morality-independent intuitions about our practical reasons, we are conceptually guaranteed that there is decisive practical reason to do whatever morality requires us to do. So even if what morality requires us to do does not correspond to any morality-independent intuitions about what there is practical reason to do, as might be true according to moral theories like Rossian Pluralism (see, e.g., Ross 1930), we have a perfectly good explanation of why we have conclusive practical reason to do these things.

I suspect that a great deal of the attraction of some moral theories, like Contractarianism, Kantianism, Contractualism, and Consequentialism, relies upon their apparent ability to explain why we should care about morality and do what morality requires by appealing to substantive morality-independent intuitions about practical reasons. I have in effect argued that these explanations of why we should be moral are misguided. As such, I believe

---

that my conceptual account of the connection between morality and practical reasons may significantly diminish the case in favor of these moral theories, and indirectly contribute to the relative attractiveness of moral theories like Rossian Pluralism, which, if we remove the desideratum that a moral theory should correspond to a substantive account of why we should be moral, may give a more plausible account of the content of morality.

References


