

# Morality, Fitting Attitudes, and Reasons for Action

Howard Nye  
University of Michigan

*Abstract:* I argue that it is actually a conceptual truth that we have reason to be moral. I defend analyses of moral concepts in terms of the fittingness of moral emotions. I argue, for instance, that we can analyze an act's moral wrongness in terms of our having reason to feel obligated not to perform it. Moral emotions like feelings of obligation involve motivations to do certain things, so the fittingness of these emotions determines the rationality of the motivations they involve. I proceed to argue that having reason to perform an act is a matter of the act's satisfying a rational motive, or contributing to an end that it is fitting to be motivated to pursue. Because morality is a matter of fitting motives, and fitting motives determine rational acts, morality entails reasons for action.

I use this strategy to explain why we have intrinsic reason to do what is moral, or reason to do so as an end in itself and quite independently of its serving other rational ends. I argue that an end's rationality is a matter of the fittingness of non-instrumental motives to pursue it, and that morality involves the fittingness of such motives. I also use my account to argue that we have conclusive reason not to do whatever is morally wrong. I contend that an act's moral wrongness entails that the reasons to feel obligated not to perform it are conclusive, which requires that there are no stronger (or equally strong) reasons to perform it.

# Morality, Fitting Attitudes, and Reasons for Action

Howard Nye

## I. Introduction

If I do something morally wrong, then absent exculpatory circumstances I should feel guilt for what I have done and others are justified in feeling angry or resentful at me for doing it.<sup>1</sup>

Moreover, it seems that if my action was genuinely morally wrong, I should not have done it in the first place. I might, in many cases quite rationally, have refrained from performing the wrongful act out of motives ranging from care for others to a desire to avoid retaliation, or simply due to a lack of any positive motivation to perform it. But absent such other motivations or tendencies to refrain from performing the wrongful action, I should have felt obligated not to perform it and should have been moved by this feeling of obligation to refrain from performing it.

I think that these kinds of platitudes are central to our moral concepts. Implicit in them is the notion of a fitting attitude, or reasons for attitudes like guilt, resentment, and feelings of obligation that contribute to their being appropriate, justified, called for, or warranted. It is important to emphasize the difference between these kinds of reasons for attitudes and what we might call pragmatic or strategic reasons to have attitudes. Were an evil demon now to threaten my loved ones with horrible punishments unless I feel guilt about having put on my left shoe before my right shoe last week, this might give me a kind of reason to get myself to feel such guilt, but it would not make the guilt fitting or appropriate.<sup>2</sup>

One's judging that an attitude is fitting seems to be capable of directly causing one to have it without one's having to do anything to get oneself to have it. Ordinarily, I do not need to

---

<sup>1</sup> As for instance Gibbard (1990) points out.

<sup>2</sup> For a similar kind of example in the case of reasons for attitudes like preferences see (Crisp 2000) and discussion in (Rabinowicz and Ronnow-Rasmussen 2004).

*try* to get myself to feel guilt if I realize that guilt would be appropriate due to my assaulting someone, breaking a promise, or telling a lie in certain circumstances – the mere realization that I should feel guilt for doing these things often seems sufficient to get me feeling guilty. On the other hand, one’s merely judging that one has pragmatic reasons to get oneself to have an attitude is not capable of directly causing one to have it – this requires further behavior or activity undertaken in order to get oneself to have the attitude. If the only reasons to feel guilt that I judge myself to have are those constituted by the demon’s threat, the only way I can form the attitude in question is for me (or some process within me) to attempt to do something in order to get me to have it (for example take a pill, condition myself, selectively attend to features of my situation or try to talk myself into thinking guilt appropriate).<sup>3</sup>

In this paper I will argue that we can analyze our moral concepts in terms of the fittingness of moral emotions. To this extent I follow A.C. Ewing (1939) and Allan Gibbard (1990). But I will also argue that we can take this kind of analysis a step further and use it to explain why exactly it is that we have the right kind of reasons to be moral. To this end I will defend a general conceptual connection between the fittingness of a motivational state and our reasons to act out of it, which I think has been more or less explicitly proposed by Brandt (1979), Bratman (1987), Gibbard (1990), Anderson (1993), Scanlon (1998), and Skorupski (1999). By

---

<sup>3</sup> While my focus here is on judgments of fittingness and pragmatic reasons for attitudes like emotions, exactly analogous remarks hold for judgments of reasons for attitudes like beliefs. What I am here calling fittingness reasons for attitudes we usually call ‘evidential reasons’ when they count in favor of beliefs, and can be contrasted with pragmatic reasons along exactly these lines. That is, judgments that one has evidential reason to believe that P can directly cause belief that P without one’s having to do anything to bring it about that one has this belief, whereas judgments that one has merely pragmatic reason to believe that P cannot cause one to believe that P without behavior undertaken to bring about one’s believing it (as has been noted by, for instance, Kavka 1983, 36).

I actually do not think that much hinges on whether we decide to call pragmatic or strategic reasons for attitudes actual “reasons for attitudes” or whether (like Gibbard (1990) and Parfit (2001)) we insist on calling them reasons to want to have or to get oneself to have attitudes, so long as we bear in mind that response to such pragmatic reasons – unlike fittingness reasons – must actually involve goal directed behavior undertaken in order to get oneself to have them. As I take this to be crucial to the distinction between fittingness and pragmatic reasons, I do in fact think that there is a principled reason to call pragmatic reasons for attitudes mere ‘reasons to get oneself to have them’. Because of this, and for ease of exposition, when in the remainder of this paper I speak without further qualification of ‘reasons for attitudes’, I intend fittingness rather than pragmatic reasons.

arguing that morality is a matter of the fittingness of certain motivational states (*viz.* moral emotions), and that the fittingness of these states entails that we have reasons to act out of them, I seek to demonstrate a conceptual connection between morality and reasons for action. I shall in fact ultimately use this strategy to attempt to defend a stronger conceptual claim in the case of moral wrongness, according to which an act's wrongness entails that one has conclusive or overriding reason not to perform it.

## II. Problems with Judgmentalism and an Alternative Approach to Moral Judgments

Perhaps the greatest initial opposition to my approach will stem from an appearance that concepts of emotions like guilt and feelings of obligation themselves involve concepts like MORAL BLAMEWORTHINESS and WRONGNESS.<sup>4</sup> I suspect that the most natural source of this appearance is attraction to a judgmentalist account of moral emotions, according to which simply feeling an emotion like guilt itself involves making a moral judgment.<sup>5</sup> I thus begin by considering some problems with judgmentalism and some virtues of my alternative approach to the relationship between moral concepts and moral emotions.

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 have feelings like 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 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

---

<sup>4</sup> My convention throughout is to use small caps to denote concepts.

<sup>5</sup> For proponents of this view see for example (Solomon 1976, 1988), (Sabini and Silver 1982), and (Foot 1959).

example, that I feel guilt for knocking over and breaking a friend's lamp, though I exercised all due caution and think 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.<sup>6</sup> 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" view according to which feeling moral emotions like guilt involve "moral evaluations" that are something less than full blown judgments, but still involve tokenings of concepts like MORAL BLAMEWORTHINESS and WRONGNESS.<sup>7</sup> Quasi-judgmentalism does not entail that episodes of recalcitrant moral emotion involve conflicting moral judgments; it entails only that they involve conflicting "moral evaluations." Why, however, should we think that recalcitrance has to involve even this?

The quasi-judgmentalist idea seems to be that moral emotions involve states that are more akin to "moral perceptions" than moral judgments.<sup>8</sup> 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 source of information, more quickly instanced, and incapable of being

---

<sup>6</sup> Cf. D'Arms and Jacobson's (2003, 129-130) 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 evidentially suspect belief is problematic."

<sup>7</sup> 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 and 129-132), and (D'Arms and Jacobson 2003).

<sup>8</sup> For instance, Roberts (1988, 187-188) 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.

consciously inferred.<sup>9</sup> 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 intuition. Return again to the guilt I feel for breaking my friend’s lamp, despite judging that I’ve 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 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.<sup>10</sup> 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

---

<sup>9</sup> See for example (Zimbardo and Weber 1997, chapter 5, especially 177-197).

<sup>10</sup> Cf. Gibbard (1990, 130): “Anyone who claims that anger includes a judgment of moral transgression needs to explain the judgment.”

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 being’s welfare), failures to adhere to certain religious practices, stringing together certain phonemes (in the form of curse words<sup>11</sup>), 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 welfare or autonomy.

Thus, a conceptual analysis of MORAL BLAMEWORTHINESS or WRONGNESS in terms of an act’s failing to maximize happiness, or being hated by deities, or violating autonomy, etc. would fail to capture our intuitions about how diverse coherent moral judgments can be and what is at issue between people with rival moral views.<sup>12</sup> 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, 126-127) gives the following such analysis of the concept of MORAL BLAMEWORTHINESS:

**Gibbard’s Analysis of Moral Blameworthiness:**

What a person does is morally blameworthy if and only if it is fitting for him to feel guilty for having done it, and fitting for others to be angry at him for having done 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.

---

<sup>11</sup> I am deeply indebted to John Ku for this example and help with this list generally.

<sup>12</sup> 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.

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 rationality of attitudes share such features as a wide diversity of things that can be coherently thought to warrant the attitudes, our attempting to determine which of these coherent positions are correct via an *a priori* method of reflective equilibrium, and our conclusions about which responses are rational 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 judgment in relation to moral emotions under those of judgments of the rationality of attitudes generally, we gain an (at least partial) explanation of them.

### **III. Attitudes and Motives**

There may, however, appear to be something missing from analyses of moral concepts in terms of the fittingness of moral emotions like guilt and anger. Moral concepts present themselves as related to reasons for *action*, or pertaining to what to *do*. For example, that something is morally wrong purports to involve our having at least some reason not to do it. Is there any way in which fitting-attitude analyses of moral concepts can explain such a connection between an action's moral status and the question of whether to perform it?



To see how they can, begin by noticing the connection between attitudes like moral emotions and motivations to act in certain ways. Feelings or emotions are not merely states of motivation; they involve qualities (for instance phenomenal and physiological) that go beyond those of mere motivations. But *part* of what it is to feel certain emotions is to be motivated to do certain kinds of things. Thus, an essential part of what it is to be afraid of, worried about, or angry at something is to be motivated, respectively, to avoid, attend to, or behave aggressively towards it.<sup>13</sup> Moral emotions in particular seem to have distinctive motivations as essential components. Guilt seems to involve something like a motivation to make amends for what one has done. Resentment and outrage or impartial anger seem to be species of anger directed towards a person, which involve motivations to behave aggressively or punitively towards him (for instance to reproach him or otherwise give him what it feels like “he has coming”).

One moral emotion that I think has not been much discussed in the recent literature is what I above referred to as a feeling of obligation. I think that we can characterize this emotion as a kind of prospective guilt-tinged aversion that one characteristically feels upon contemplating the prospect of performing or failing to perform an action, which performance or omission one takes to constitute doing something morally wrong. I think that Brandt (1959) gives a nice description of this emotion in his discussion of his “parked car episode”:

[I noticed] a car pulled off to the side of the road, with a man in it slumped over the wheel – as if asleep or ill... There was actually an impulse to stop.... In such cases, we normally say that we did not want to do a certain thing, but did it because we thought we ought. The writer was... not considering stopping in order to terminate some organic discomfort... Nor... in order to get or do something for himself... Some psychologists, at least partly in order to give recognition to the distinctiveness of the motivation, have suggested calling this sort of experience an “experience of requiredness”; but... there is no reason why we should not use an ordinary mode of speech to cover it: “I felt an obligation to...” (Brandt 1959, 116-118).

---

<sup>13</sup> We seem, moreover, to be able to distinguish between on the full-on feelings of emotions like the fear we feel towards disasters we perceive to be actually impending, and imaginatively simulated such feelings or feelings in response to imagined scenarios, like impending disasters portrayed in the fiction of a movie. Unlike the former, the latter kind of emotion or simulation of emotion involves no real motivation to avoid or flee something we take to be actual – at most it involves a motivation to avoid or flee something in our imaginative simulation of a situation.

This feeling of obligation or prospective guilt-tinged aversion is also what Mill (1863, Chapter III, paragraphs 3 and 4) seemed to describe 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.”

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 obligated to stay with him. As with our discussion of recalcitrant guilt in Section II, it seems that the man and woman could in this way recalcitrantly feel obligated without any of the conflicting inferential tendencies, views about appropriate conduct, or views about their own responses required for conflicting judgments about what they are morally obligated to do. Moreover, it seems 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.

Feelings of obligation or prospective guilt-tinged aversion thus seem explicable, not as states that essentially involve tokenings of MORAL OBLIGATION or WRONGNESS, but rather as manifestations of a phenomenal and motivational syndrome that – among other things – involves motivation to perform (or refrain from performing) the actions towards which they are felt. It is

very important to distinguish these motivations involved as components of such feelings of obligation from motivations to do things in order to avoid feeling guilt or guilt-tinged aversion. If one could take a pill that would prevent all unpleasant feelings of guilt for doing something, or indeed a pill that would terminate all uncomfortable feelings of obligation not to do it, the latter kind of motivation not to do it in order to prevent or terminate the feelings of guilt or obligation would be eliminated in favor of a motivation to take the pill. But the motivation not to do things involved as a component of feeling obligated not to do them would not be affected by the availability of such pills. When one feels such prospective guilt-tinged aversion towards doing A, one either feels this aversion non-instrumentally towards doing A (for instance A = harming another being), or towards A because one believes it to be a sufficient means or constitutive way of doing something else B, towards which one feels non-instrumental prospective guilt-tinged aversion (for instance A = pulling the trigger on a fire-arm aimed at a being and B = harming the being).

#### **IV. An Analysis of Moral Wrongness and An Account of Reasons not to do Wrong**

I have so far considered how in addition to retrospective guilt we have an emotion of feeling obligated to do something, and how such feelings involve motivations to perform the actions towards which they are felt. But how exactly should this prospective guilt-tinged aversion to failing to do something, and our reasons to feel it, figure into our understanding of concepts like MORAL WRONGNESS and BLAMEWORTHINESS? Above I discussed how Gibbard seeks to analyze moral blameworthiness in terms of the fittingness of anger and *retrospective guilt*. But Gibbard crucially draws attention to some ways in which the concepts of moral blameworthiness and moral wrongness can come apart, including the following example:

Imagine that in a paroxysm of grief I speak rudely to a friend who offers condolences, and so hurt his feelings. My rudeness is unprovoked, but understandable in the circumstances. I have thus acted wrongly, but because of my agitated state, it may not make sense to blame me (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*. Unlike blameworthiness, assessments of moral wrongness are most intimately related to asking which of the acts open to oneself are wrong, and being motivated not to perform these. 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 with the rationality of prospective guilt-tinged aversion rather than retrospective guilt. Indeed, the same reasons that tell in favor of Gibbard’s analysis of MORAL BLAMEWORTHINESS in terms of the rationality of retrospective guilt and impartial anger support the following analysis of MORAL WRONGNESS in terms of the rationality of prospective guilt-tinged aversion:

**The Quasi-Gibbard Analysis of Moral Wrongness:**<sup>14</sup>

Agent S’s doing A is morally wrong if and only if it is fitting for S to feel prospective guilt-tinged aversion towards doing A (or equivalently fitting for S to feel obligated not to do A).

Like Gibbard’s analysis of moral blameworthiness, this analysis has the virtue of explaining what is common to all coherent judgments about moral wrongness, which we have seen can attribute wrongness to wildly disparate acts. But this analysis can also help explain the gap

---

<sup>14</sup> I say “Quasi-Gibbard” because Gibbard himself does not give this analysis explicitly, but in the spirit of his approach it seeks to analyze a moral concept in terms of the rationality of a moral emotion, and I think something like it was what Gibbard was trying to get at in his explicit account of moral wrongness in (Gibbard 1990). Gibbard (2006, 2007) also seems to take a favorable attitude towards this analysis, or at least the use of prospective guilt-tinged aversion in the analysis of moral concepts.

Gibbard noted between judging an act wrong and judging it blameworthy. Combining it 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 rational for us to feel angry at the person who lashed out and it isn't rational 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.<sup>15</sup>

A final virtue of this analysis is that, given the influence that judgments of (fittingness) reasons for attitudes exert on our actually having them, this analysis can explain how moral judgments can be motivating – not only in the sense that they can motivate our trying to make amends for blameworthy wrongs we take ourselves to have committed, but how they can moreover motivate us to refrain from committing what we take to be wrongs in the first place. Since prospective guilt-tinged aversion towards performing an act involves motivation not to

---

<sup>15</sup> One might be wondering, however, why there is not in addition to a wrongness-blameworthiness gap a blameworthiness-wrongness gap, or why it is that although an act can warrant prospective guilt-tinged aversion before the fact but not outrage or guilt after the fact, it can never warrant outrage and guilt after the fact but not prospective guilt-tinged aversion before the fact. One thing I should point out is that conceptual connections between warrants for distinct moral emotions is already an issue for the Gibbardian analyses, in that their proponents need to explain why acts cannot warrant outrage but not guilt or guilt but not outrage. 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 warrants were interrelated in these ways.

Very briefly, here is why I think this came to be the case. For evolutionary reasons, our ancestors tended to feel guilt and outrage after the fact only towards those acts towards which the performer felt prospective guilt-tinged aversion before the fact (but not vice versa). 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 after the fact only towards those acts towards which they also prescribed feeling prospective guilt-tinged aversion before the fact (but not vice versa). 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 prospective guilt-tinged aversion 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 or prospective guilt-tinged aversion were extracted via the Ramsey-Carnap-Lewis method of analyzing theoretical concepts (see for instance (Lewis 1970, 1972)).

perform it, judging that one should feel this guilt-tinged aversion towards performing an act will tend to cause one to actually feel the aversion, including its motivation not to perform the act.<sup>16</sup>

Suppose, then, that an act – say lying in a particular situation – is morally wrong. According to the Quasi-Gibbard analysis, this means that one should feel prospective guilt-tinged aversion towards lying in that situation. What follows? As we have seen, prospective guilt-tinged aversion to doing something has as an essential component a motivation not to do it. But if it is rational for me to feel prospective guilt-tinged aversion towards lying in a situation, and this involves being motivated not to lie, is this motivation not also rational? And if this motivation not to lie is rational, do I not also have reason not to lie?

To make matters more precise, begin by considering the following principle, which I think should readily be admitted as a conceptual truth about reasons or warrant for attitudes:

**Warrant composition principle [WCP]:**

Let  $\phi$  be a psychic state that involves  $\psi$  as an essential component. If we have fittingness reason to be in  $\phi$ , then we have fittingness reason to be in  $\psi$ .

WCP simply states that if one has reason to be in a psychic state, one necessarily has reason to be in all that the state essentially amounts to. Since feeling prospective guilt-tinged aversion to doing something essentially involves motivation to refrain from doing it, WCP entails that if one

---

<sup>16</sup> Gibbard (1990) gives an impressive expressivist semantics of judgments of warrant or fittingness reasons for attitudes that seeks to explain how such judgments have this kind of attitude guiding feature. One of the central initial motivations for expressivism was an attempt to explain how moral judgments can be motivating, and paired with his analysis of blameworthiness, Gibbard's expressivist semantics can admirably explain how judgments of blameworthiness can motivate making amends and punishing transgressors. Unfortunately Gibbard's expressivist semantics cannot combine with either his analysis of blameworthiness or the explicit analysis of wrongness in (Gibbard 1990) to similarly explain what is perhaps the most central case of moral motivation – that of an actor not to perform actions she deems to be wrong. But Gibbard can immediately remedy this defect by adopting the Quasi-Gibbard analysis of wrongness, enabling the central case of motivation not to do what one judges to be wrong to be explained along the same lines as motivation to make amends for and punish what one judges to be blameworthy.

has reason to feel prospective guilt-tinged aversion towards doing something, one has reason to be motivated not to do it.

Now, as I suggested it seems that the fact that prospective guilt-tinged aversion is rational makes it rational not only to have the motivation it involves but indeed to act out of it. The general idea here seems to be that if one has reason to be motivated in a certain way, one also has reason to act out of that motive. It might appear, however, that not all reasons to have motives translate into reasons to act out of them. For instance, in a variant of Kavka's (1983) Toxin Puzzle, one will receive \$1,000,000 if a completely reliable motivation detector senses that one is motivated to take a toxin that will make one sick for a day.<sup>17</sup> Here it seems that in a sense one has reason to have a motivation to take the toxin, but surely this does not mean that one has reason to act out of the motive and actually take it (after all one is paid only for having the motivation; actually taking the toxin means getting sick needlessly). But recall the distinction between fittingness reasons and pragmatic or strategic reasons to have attitudes discussed in Section I. In the toxin puzzle, one's "reasons to have a motive to take the toxin" are merely pragmatic reasons to get oneself to have the motive, rather than fittingness reasons or considerations that contribute to the motive being appropriate or warranted. With this distinction in mind, I contend that the following is also a general conceptual truth about the relationship between reasons for motivation and reasons for action:

---

<sup>17</sup> Kavka's original puzzle is told with being offered \$1,000,000 to form an intention to take the toxin, but I think that my points here do not depend upon the relationship between intentions and other motivational states.

**Motivation-Action principle [MAP]:**

Let  $\phi$  be an action. If one has fittingness reason to be motivated to  $\phi$ , then one has reason to actually  $\phi$ .<sup>18</sup>

The idea behind MAP is that since it is through the rational assessment of our motivations that our actions are guided by reason, our reasons for action *just are* reasons for motivations to perform these actions. To see this, start by asking: what makes an action rational or something we really should be doing? Some have claimed that an agent's action is rational or what she should do just in case her performing it will bring about her actual ends, or the ends she is actually non-instrumentally motivated to bring about. But surely this is wrong – when our actual non-instrumental motivations are irrational, it is equally irrational to act out of them and do things to bring about their objects. For instance, if out of an irrational kind of esteem for a stranger and an irrational lack of care for my child I non-instrumentally care more for the welfare of this stranger than that of my child, it seems that no matter how strong this irrational non-instrumental preference, I have no reason to act out of it.

But it does seem that the *rationality* of an end or a non-instrumental motive translates into the rationality of actions out of it or actions that bring about its object. As I have *reason* to non-instrumentally care more for my child than a stranger I don't have much reason to esteem, I

---

<sup>18</sup> My warrant composition and motivation-action principles closely correspond to Skorupski's partition of his "Feeling/Disposition Principle," (FD): "if there is reason to feel  $\phi$  then there's reason to do that which  $\phi$  disposes to" into (FDF): "If there's reason to feel  $\phi$  there's reason to desire to do that which  $\phi$  characteristically disposes one to desire to do," and (FDD): "If there's reason to desire to do  $\alpha$  (or to bring it about that  $p$ ), there's reason to  $\alpha$  (to do that which will bring it about that  $p$ )" (see (Skorupski 1999), especially p. 38, 63, 131, and 174 n24). The main possible differences between my principles and Skorupski's FD principles is my WCP's insistence on  $\psi$ 's essential involvement as component of  $\phi$  rather than (in his FDF) its mere involvement as what " $\phi$  characteristically disposes one to desire to do" and my MAP's clarification that the relationship between reasons for motivations and actions is held to hold only between fittingness (and not pragmatic) reasons for motivation and reasons to act out of them.

Here I am trying to give general rationales for my warrant composition and motivation-action principles, which I would hope would largely vindicate Skorupski's (FDF) and (FDD) when suitably interpreted. Another main difference between Skorupski and myself is, of course, that I am pairing them with my Quasi-Gibbard analysis of moral wrongness in an attempt to explain our reasons not to do wrong.



do have reason to act out of such a non-instrumental preference. It seems, then, that the rational assessment of an agent's actions is dependent, not on her actual ends or non-instrumental motives, but on which non-instrumental motives it would be rational for her to have.

What the lines of thought behind WCP and MAP would seem to show us, then, is that deliberation about which ends are rational is possible and proceeds via deliberation about which non-instrumental motivational states (including emotions but also desires in the ordinary English sense) are warranted, fitting, or appropriate. To think an ultimate end rational is to think it rational to be non-instrumentally motivated to pursue it. Now, as we saw above, feelings of prospective guilt-tinged aversion towards doing something are either (i) non-instrumental or (ii) towards the act as a sufficient means or constitutive way of doing something else towards which one feels prospective guilt-tinged aversion non-instrumentally. To judge an act to warrant prospective guilt-tinged aversion is thus either to judge it to warrant non-instrumental guilt-tinged aversion or guilt-tinged aversion as a sufficient means or constitutive way of doing something else that warrants non-instrumental guilt-tinged aversion. With the Quasi-Gibbard analysis we can understand the former as judgments that the act is intrinsically morally wrong and the latter as judgments that it is instrumentally morally wrong. This Quasi-Gibbard Analysis, in conjunction with our account of judging an ultimate end to be rational, entails that if an act is intrinsically wrong then avoiding doing it is a rational ultimate end. This analysis and account also entail that if an act is instrumentally wrong, we have reason not to perform it simply as a means to the rational end of avoiding doing the intrinsically wrong thing it is a way of doing.

Suppose, for example, that harming an innocent being in a particular situation is intrinsically wrong. According to the Quasi-Gibbard analysis, it is a conceptual truth that this amounts to its being fitting in such a situation to feel non-instrumental prospective guilt-tinged

aversion to harming the being. By the line of thinking behind WCP and MAP, this analytically entails that avoiding harming the being in such situations is a rational ultimate end. In thinking it intrinsically wrong to perform the act, we are already committed to thinking it rational to be non-instrumentally motivated not to perform it, and thus that we have reason not to perform it as an end in itself. Similarly, our thinking that it is instrumentally wrong to perform an act like pulling the trigger of a gun aimed at the being already commits us to thinking that we have reason to avoid performing it simply as a means to the rational end of avoiding the intrinsically wrong conduct - here harming the being – that it would bring about.

If this account is correct, then it should be clear how we have reason not to perform those acts that are genuinely morally wrong without our having to show that refraining from performing them is conducive to some other rational end, like our own welfare or an overall better state of the world. This, I think, explains how we have the kind of “intrinsic” or “right kind of reasons” to avoid doing what is morally wrong that people like H.A. Prichard (1912) and W.D. Falk (1948, esp. 23) have been concerned with. Moreover, if my account is correct these reasons follow from an act’s moral wrongness as a conceptual as opposed to simply a substantive normative matter. Whether or not any particular act is in fact morally wrong is of course a substantive normative question, but no further substantive step is needed to take us from the fact that it is morally wrong to the fact that we have reason not to perform it, either as an end in itself or as a way of avoiding doing something else that is intrinsically morally wrong.

## **V. Moral Reasons in General**

So far, I have argued we can analyze concepts like MORAL BLAMEWORTHINESS and WRONGNESS in terms of the fittingness of moral emotions, and that we can explain the intuitive connection

between an act's moral wrongness and our reasons not to perform it in terms of a general conceptual connection between fitting motivational states and reasons to act out of them. While it is beyond the scope of this paper to discuss the details, I think that we can apply this same strategy of analysis to explain other moral concepts, like those of MORAL GOODNESS (or MORAL ESTIMABILITY) and MORAL BADNESS (or MORAL DISESTIMABILITY), and their intuitive connections to reasons for action as well. For instance, I think that Brandt's (1946, 113) general suggestion that "'X is Y-able'...means that 'X is a fitting object of Y-attitude (or emotion)'" supports the following analyses:

**Brandtian Analyses of Moral Goodness and Badness:**

Let  $\Phi$  be an act motivated in a particular way, a motive or motivational state (like a particular desire or emotion), a character trait, or an agent.

$\Phi$  is morally good or estimable if and only if it is fitting to morally esteem  $\Phi$ , and

$\Phi$  is morally bad or disestimable if and only if it is fitting to morally disesteem  $\Phi$ .

Morally esteeming something feels like "looking up to" its object, feels akin to an impartial version of gratitude or thanks, and involves wishfully imagining or fantasizing about doing or being like its object.<sup>19</sup> Importantly it involves motivation to emulate its object - to perform acts esteemed or behave like the agent or one with the trait or motive esteemed. In contrast, morally disesteeming something feels like "looking down on" its object, is akin to an impartial version of feeling "screwed over," and involves wishfully imagining or fantasizing about doing or being *unlike* its object in similar circumstances. Importantly it involves motivation to *disemulate* its object - to refrain from performing acts disesteemed or to behave unlike the agent or one with the

---

<sup>19</sup> Cf. (Velleman 2002, esp 101).

trait or motive disesteemed. Since moral esteem and disesteem involve these motivations, it follows from the Brandtian analyses and the warrant composition and motivation-action principles that we have reason to perform morally good acts and act like the morally good, as well as reason not to perform morally bad acts and reason not to act like the morally bad.

What this suggests, I think, is that we can apply this strategy to *all* moral concepts, analyzing them in terms of fitting moral emotions and using the warrant composition and motivation-action principles to explain their conceptual connections to reasons for action.<sup>20</sup> If this is so, then we can understand a moral reason for or against performing an action as a consideration that contributes to that action's instantiating a moral concept, but that *ipso facto* also counts in favor of feeling the moral emotions the fittingness of which constitutes its instantiating the concept, having the motivations these emotions involve, and performing or refraining from performing the action out of these motives.

## **VI. 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 falling under a moral concept and our having reason to (or not to) perform it. Instantiations of certain moral concepts seem to entail that we have some reason to do things, but not conclusive reason. If an act is morally good we have some reason to perform it, but it at least seems coherent to think that there are morally

---

<sup>20</sup> One might be wondering what unifies the moral emotions and distinguishes them from other emotions (like for instance shame, contempt, care, scorn, fear, etc.). The answer, I think, is that they differ phenomenally from other emotions and it is a conceptual truth about them that there are entailment relations between their warrants (for instance if an act warrants outrage, guilt, or resentment after the fact, then it also warrants prospective guilt-tinged aversion before the fact and all three of outrage, guilt, and resentment after the fact). As I mentioned in note 15, I believe that we came to have emotion concepts with these entailment relations by ramsifying over a folk psychological theory that was true of us and the truth of which involved, for evolutionary reasons, roles for emotion states that were jointly governed by norms.

good actions that it can be rationally permissible to fail to perform, like getting oneself shot in the head to spare a stranger the somewhat more painful death of drowning.

But the notion of MORAL WRONGNESS OR MORAL OBLIGATION (where what's morally obligatory is just what it's morally wrong not to do<sup>21</sup>) 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 most 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 obligation, namely "if S is morally obligated 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 obligatory or wrong. 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<sup>22</sup>, 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*.<sup>23</sup> 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 OBLIGATION – unlike, say, our concept of MORAL GOODNESS – is itself

---

<sup>21</sup> This is at least one clear sense of the term 'moral obligation' in English. Perhaps one can also speak in English of 'moral obligations' in a "prima-facie" (or really a *pro tanto*) sense, or in a way in which one can have one "moral obligation" to  $\phi$  without its being morally wrong for one to fail to  $\phi$  if one has a "stronger moral obligation" not to  $\phi$ . If so, then I merely wish to distinguish the concept I am expressing with 'moral obligation' from this other concept, and suggest that the concept expressed by 'moral obligation to  $\phi$ ' in this other sense is what I would call the notion of THERE BEING CONSIDERATIONS THAT CONTRIBUTE TO ONE'S BEING MORALLY OBLIGATED TO  $\phi$  (OR THE MORAL WRONGNESS OF ONE'S FAILING TO  $\phi$ ).

<sup>22</sup> 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.

<sup>23</sup> 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: (i) were it not for the harm to oneself one would be morally obligated to keep the promise, but (ii) given the harm one would incur by keeping the promise it is rationally permissible to break the it, yet (iii) one does not "owe it to oneself" to prevent the harm to oneself by breaking the promise.

sensitive to whether or not the reasons in favor of performing it are actually conclusive. That is, as W.D. Falk (1948, 30-31) suggested, “our very thinking that we ought [that is, are morally obligated] 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 quasi-Gibbard 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 blameworthy. That an act is blameworthy does not entail that others are necessarily irrational if they fail to feel angry with the blameworthy actor. Especially if the transgression is slight, others might be rationally permitted not 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.<sup>24</sup> 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* reason to feel obligated not to perform it.<sup>25</sup>

---

<sup>24</sup> Cf. (Gibbard 1990, 126-127).

<sup>25</sup> 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.

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 be a reason in favor of wanting to eat them. That the foods are unhealthy seems to be a reason in favor of being averse to eating them, and refraining from eating them out of this aversion, but intuitively it does not seem to be a reason against wanting 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 we can in such cases have conclusive reason to (i) be somewhat motivated to eat the foods, (ii) be more strongly motivated not to eat them, and (iii) act out of our motives not to eat them and refrain from doing so.

What this kind of case suggests is that we need to distinguish between having conclusive reason to have *some* motivation to perform an act and having conclusive 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, is that while one can have conclusive reason to be gradationally motivated to do something without having conclusive

reason to do it, one's having conclusive 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 reason to be most motivated to do A, then S has conclusive reason to do A.

We have thus seen how for some gradational motivations, reasons to be in conflicting motivational states do not themselves constitute 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. Reasons to be motivated not to do something actually *do* seem to count against 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 reason, not only in favor of being most motivated not to keep one's promise, but indeed a reason against having any feeling of obligation to keep it under the circumstances.<sup>26</sup>

---

<sup>26</sup> 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, 742-743), 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 the home to think that there is something wrong with one if one does not feel prospective guilt-tinged aversion towards 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 inappropriate or 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 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 such aversion to putting her in the home. Given my discussion in Section IV, I think we should understand this as a judgment that it is fitting to feel moral disesteem towards not feeling such prospective guilt-tinged aversion. Such a judgment may easily be mistaken for a judgment that it is actually fitting to feel prospective guilt-tinged aversion because the former



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 reason to feel obligated to do *A*, then *S* has conclusive reason to be most motivated do *A*.<sup>27</sup>

---

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 prospective guilt-tinged aversion towards 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 or prospective guilt-tinged aversion the 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 or prospective guilt-tinged aversion to performing an act do not seem so aptly characterized in these ways – they seem to involve something more like a feeling that one "just can't" or "just can't bring oneself" to perform the act. Compunction seems to be more closely associated with going back and forth about or checking and re-checking to make sure about what it would be wrong for one to do.

<sup>27</sup> 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 or prospective guilt-tinged aversion 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 or prospective guilt-tinged aversion conclusive reason for which constitutes the moral obligatoriness or moral wrongness of 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 footnotes 15 and 20. 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 guilt-tinged aversion only when they required no stronger (or equally strong) motives to the contrary. In this way the folk psychological theory that came to be true of us was one where the states that played the prospective guilt-tinged aversion 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 or PROSPECTIVE GUILT-TINGED AVERSION.

We can now combine these considerations in favor of the most-motivation-action principle and the contour thesis with the Quasi-Gibbard analysis of moral wrongness to vindicate the strong morality-reasons internalism thesis about moral obligation. Given our clarification of the Quasi-Gibbard 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 reason to feel obligated to do A. If S is already sufficiently moved to do A, then either S is in a rationally permissible state of being most motivated to do A, or S's state of being most motivated to do A is rationally impermissible. 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 reason to feel obligated to do A. By the contour thesis, if S has conclusive reason to feel obligated to do A, then S has conclusive reason to be most motivated to do A. Thus, if it would be morally wrong for S to fail to do A, reason will only allow S to be most motivated to do A, so S has conclusive reason to be so motivated. Finally, by the most-motivation-action principle, if S has conclusive 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 obligated to do A, then S has conclusive reason to do A.<sup>28</sup>

---

<sup>28</sup> I should point out that if one doubts the truth of the contour thesis for reasons related to cases like that discussed in footnote 26, there are still at least two other ways in which my approach to moral concepts and their connection to reasons for action can help 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

It may be important to conclude by emphasizing what this kind of vindication of the strong morality-reasons internalism thesis does and does not show. On my account, this thesis is true only because: (1) for S to be morally obligated to do A, it must be the case that absent a sufficient motivation or tendency to do A anyway S has conclusive reason to feel obligated to do A, and (2) in contrast to reasons for some other motivational states, if the 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 reason to have the feeling of obligation at all. Like other kinds of reasons, an agent's reasons to 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 the agent should feel and hence (given the Quasi-Gibbard analysis of moral wrongness) are no longer sufficient to make it morally wrong for the agent to do otherwise, or morally obligatory 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 obligation-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 obligated to do something, then we have conclusive reason to do it. I think that our evidence that we are morally obligated to do things is just as good as our evidence that other normative concepts are instantiated, and that this consists in the best unification and explanation of our normative intuitions by normative theories in reflective-equilibrium. To the extent that our best

---

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.

such normative theories tell us that we are morally obligated to do something, I think we should conclude that we are, which I have argued entails that we have conclusive reason to do it. This is no more than I think most of us have intuitively suspected all along. But if my confirmation of our suspicion is correct, let us not forget that it is so, even if we are sometimes tempted to do so when faced with evidence that we are morally obligated to do inconvenient or difficult things.

## REFERENCES

- Anderson, Elizabeth. 1993. *Value in Ethics and Economics*. Cambridge: Harvard University Press, 1993.
- Brandt, Richard B. 1946. Moral Valuation. *Ethics*, 56: 106-121.
- Brandt, Richard B. 1959. *Ethical Theory*. Englewood Cliffs, N.J: Prentice Hall.
- Brandt, Richard B. 1979. *A Theory of the Good and the Right*. Amherst, New York: Prometheus Books.
- Bratman, Michael E. 1987. *Intention, Plans, and Practical Reason*. Cambridge, Massachusetts: Harvard University Press.
- Crisp, Roger. 2000. Value... and What Follows by Joel Kupperman. *Philosophy* 75: 458-462.
- D'Arms, Justin and Daniel Jacobson. 1994. Expressivism, Morality, and the Emotions. *Ethics*, 104: 739-765.
- D'Arms, Justin and Daniel Jacobson. 2003. The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism). *Philosophy: The Journal of the Royal Institute of Philosophy*, 52 (suppl.): 127-145.
- Darwall, Stephen L. 1997. Reasons, Motives, and the Demands of Morality. In *Moral Discourse and Practice*, edited by Stephen Darwall, Allan Gibbard, and Peter Railton, 305-312. New York: Oxford University Press.
- Ewing, A.C. 1939. A Suggested Non-Naturalistic Analysis of Good. *Mind*, 48: 1-22.
- Falk, W.D. 1948. 'Ought' and Motivation. *Proceedings of the Aristotelian Society*, 48: 111-138. Reprinted in *Ought, Reasons, and Morality: the Collected papers of W.D. Falk*, by W.D. Falk. Ithica: Cornell University Press.
- Foot, Philippa. 1959. Moral Beliefs. *Proceedings of the Aristotelian Society*, 59:83-104.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, Massachusetts: Harvard University Press.
- Gibbard, Allan. 2006. Reply to Critics. *Philosophy and Phenomenological Research*. 72: 729 - 744.
- Gibbard, Allan. 2007. Thinking How to Live Together. In Grethe B. Peterson, ed., *The Tanner Lectures on Human Values*, vol. 27 (Salt Lake City: University of Utah Press), 165–226.

- Greenspan, Patricia. 1988. *Emotions and Reason: An Inquiry into Emotional Justification*. London: Routledge & Kegan Paul.
- Kavka, Gregory S. 1983. The Toxin Puzzle. *Analysis*, 43: 33-36.
- Kitcher, Philip S. 1998. Psychological Altruism, Evolutionary Origins, and Moral Rules. *Philosophical Studies*, 89: 283-316.
- Lewis, David. 1970. How to Define Theoretical Terms. *Journal of Philosophy*, 67: 427-446.
- Lewis, David. 1972. Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, 50: 249-58.
- Mill, John Stuart. 1863. *Utilitarianism*. London: Parker, Son, and Bourn.
- Parfit, Derek. 2001. Rationality and Reasons. In *Exploring Practical Philosophy: From Action to Values*, edited by Dan Egonsson, Bjorn Petersson, Jonas Josefsson, and Toni Ronnow-Rasmussen, 17-41. Aldershot: Ashgate.
- Prichard, H.A. 1912. Does Moral Philosophy Rest on a Mistake? *Mind*, 21: 21-37
- Rabinowicz, Wlodek and Toni Ronnow-Rasmussen. 2004. The Strike of the Demon: On Fitting Pro-attitudes and Value. *Ethics* 114: 391-423.
- Roberts, Robert C. 1988. What an Emotion Is: A Sketch. *The Philosophical Review*, 97: 183-209.
- Sabini, John and Maury Silver. 1982. *Moralities of Everyday Life*. Oxford: Oxford University Press.
- Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, Massachusetts: Harvard University Press
- Skorupski, John. 1999. *Ethical Explorations*. New York: Oxford University Press.
- Solomon, Robert C. 1976. *The Passions*. Garden City, N.Y.: Doubleday/Anchor.
- Solomon, Robert C. 1988. On Emotions as Judgments. *American Philosophical Quarterly*, 25: 183-191.
- Velleman, J. David. 2002. Motivation by Ideal. *Philosophical Explorations*, 5: 90-104.
- Zimbardo, Philip G. and Ann L. Weber. 1997. *Psychology*. 2<sup>nd</sup> Edition. New York: Longman.