The Weight of Moral Reasons
Ralph Wedgwood
University of Southern California

Many thinkers have held that one of the central issues of moral philosophy is a question about the reasons – if any – that we have for conforming to moral requirements. In effect, this is the question “Why be moral?” Many different answers to this question have been attempted – including Humean, Hobbesian, and Kantian answers, among others.

This essay aims to develop a different sort of answer – an answer that is situated within the framework of a broadly value-based conception of reasons for action, of roughly the sort that has been advocated by Joseph Raz (1999a). According to such value-based conceptions, every reason for action corresponds to a fact about how the available options instantiate some appropriate value. Roughly, according to these conceptions, you have a reason to φ if and only if the option of φ-ing is available to you, and is in an appropriate way a good thing to do.

First, however, we have to clarify what exactly our central question means. This will help us to understand what would count as an adequate way of answering it.

1. What is the question?

Not all moral philosophers accept that there is a good question to answer here. For example, the 18th-century British moralist Richard Price (1787, 180) writes:

To ask, why are we obliged to practise virtue, to abstain from what is wicked, or perform what is just, is the very same as to ask, why we are obliged to do what we are obliged to do? — It is not possible to avoid wondering at those, who have so unaccountably embarrassed themselves, on a subject that one would think was attended with no difficulty.

In view of how many philosophers have been seriously perplexed by this question, Price’s interpretation of the question is too uncharitable to be credible. In what follows, I shall interpret the question in the following way. First, let me introduce the idea of a “non-trivial moral requirement”. Roughly, a requirement that you are subject to counts as non-trivial if and only if it is possible for you not to conform to it, and it would intelligible for you to be tempted not to. (I shall say more about such “non-trivial” requirements later.)

---

1 For a more detailed version of this approach to reasons for action, see Wedgwood (2009a).
With this notion of non-trivial requirements in hand, we can now interpret our central question as concerned with the following two propositions:

1. All normal human beings are subject to non-trivial moral requirements.
2. Whenever any agent is subject to a non-trivial moral requirement, that agent has overriding or conclusive reasons for conforming to it.

Our central question is, in effect: What explains why these two propositions are true?

Some philosophers might be tempted to agree with H. A. Prichard (1912, 21) that this question is “improper” – on the grounds that these two propositions are simply utterly primitive truths, which cannot be explained on the basis of any other truths whatsoever. However, it is not clear that there is any compelling argument for the conclusion that these propositions are primitive truths of this sort.

One famous point that Prichard (1912, 23) made in the course of arguing for this conclusion is that it cannot be that the only conclusive reason that we have for conforming to moral requirements is that it is in our self-interest to do so. In what follows, I shall accept this point: according to the account that I shall propose here, the overriding or conclusive reasons that we have for conforming to moral requirements must always include distinctively moral reasons, and need not include any reasons of self-interest at all. However, this point is obviously compatible with the thought that the correct account of the nature of reasons for action and of moral requirements could help us to understand why there are any moral reasons at all, and why these moral reasons include overriding or conclusive reasons for conforming to all moral requirements.

This is the sort of answer to our central question that I shall search for here. I shall start by giving a sketch of a general conception of reasons for action and of moral requirements; and then I shall argue that this conception can help us to understand why these two propositions are true.

2. A general conception of reasons for action

First, we need a conception of reasons for action. I shall start by making some comments on what it is for the reasons that you have to φ at t to count as “conclusive” or “overriding” reasons.2

Overriding or conclusive reasons seem to satisfy the following necessary condition. If an agent’s reasons for φ-ing at a time t count as overriding or conclusive, then the agent ought, all things considered, to φ – where this occurrence of ‘ought’ is what we could call the “objective practical ‘ought’”, indexed to the situation of this agent at this time t. The practical ‘ought’, as I

---

2 Some philosophers, such as Robert Audi (1993), define different senses for the two expressions ‘You have a reason to φ’ and ‘There is a reason for you to φ’. As I use these two expressions, however, they are synonymous and have exactly the same sense.
understand it, is neither moral nor non-moral, but reflects all reasons or considerations that bear on how the agent should act at \( t \); it is, in other words, an all-things-considered ‘ought’. The “objective” practical ‘ought’ differs from more “subjective” (or “information-relative”) versions of the practical ‘ought’ in so far as the facts that determine how the agent ought to act or choose at \( t \) are not limited to facts that the agent knows or is even in a position to know at \( t \).

Indeed, it may be that the following more detailed connection holds between the objective practical ‘ought’ and the notion of what we have overriding or conclusive reasons to do: it may be that the agent ought (in this sense) to \( \varphi \) at this time \( t \) if and only if \( \varphi \)-ing is at least part of the overall course of action that the agent has overriding or conclusive reasons to take at \( t \). For our purposes, however, it does not matter whether this more detailed connection is exactly right. What we need is simply the necessary condition that I articulated above – that is, the thesis that if an agent has overriding reasons to \( \varphi \) at \( t \), then the agent ought (in the objective all-things-considered practical sense) to \( \varphi \) at \( t \).

This thesis makes it plausible that there is a more general connection between reasons and the practical ‘ought’. At least so long as we restrict our attention to cases where it is possible for the relevant agent not to \( \varphi \), it seems that if the agent ought to \( \varphi \), there must be some explanation or reason why the agent ought to \( \varphi \) – and this explanation of why the agent ought to \( \varphi \) can surely be called a reason for the agent to \( \varphi \).

In general, then, it seems plausible to follow the seminal suggestion of John Broome (2004), according to which a reason for an agent \( x \) to \( \varphi \) at a time \( t \) is a fact that plays what we could call the “pro \( \varphi \)-ing” role in an explanation of how \( x \) ought to act at \( t \). Indeed, it seems plausible to me to suggest that all the facts about how an agent ought to act at \( t \) are explained by the following two factors:

\( a. \) Facts about what options are available to the agent at the time \( t \);

\( b. \) The reasons for and against each of those options.

What is it exactly for a fact to play the “pro \( \varphi \)-ing” role in an explanation of how \( x \) ought to act at \( t \)? Again, I cannot attempt a full answer to this question, but I will try to clarify what is at issue by articulating some necessary conditions on reasons for action. As I mentioned at the outset, my goal here is to explore the implications of a distinctively value-based view of reasons for action. So I shall assume here that all reasons for action consist in facts about how the available options exemplify various values – i.e., facts about how these options are, in various ways, good or bad, or better or worse than each other.

More specifically, the conception of reasons for action that I shall assume here implies that a fact counts as a reason for you to \( \varphi \) if and only if that fact entails that \( \varphi \)-ing is in an appropriate way a good thing for you to do. In addition, I shall also assume here that goodness is reducible to

\[ 3 \text{ For a more detailed account of the practical ‘ought’, see Wedgwood (2007, chapter 4).} \]
betterness. That is, for a course of action to be good is simply for it to be better than the appropriate benchmark of comparison. I shall not be able to say anything about this “benchmark of comparison” here, but presumably, if a course of action is better than this “benchmark”, then that course of action must at least be better than some of the available alternatives.

There are of course many different ways in which a course of action can be as good as or better than some alternative. It is doubtful whether all of these different ways of being good can generate genuine reasons for action. If not, then only some of these ways of being good can ground any genuine reasons – in which case a full account of reasons for action would have to include a more precise account of which such ways of being good can do this. For present purposes, however, what has been said so far will suffice to give us a conception of reasons to work with in the rest of this discussion.

3. A general conception of moral requirements

What are “moral requirements”? It seems plausible to me that ‘required’ here means neither more nor less than ‘needed’, which seems to refer to what is conditionally necessary – in this case, what is necessary for avoiding moral wrongness. More precisely, then, we can give the following definition of what it is for an agent to be morally required to φ at a given time $t$:

An agent $x$ is morally required to $φ$ at a time $t$ if and only if, in every state of affairs available to $x$ at $t$ in which neither $x$’s practical reasoning nor $x$’s behaviour at $t$ is morally wrong, $x$ $φ$-s.\(^5\)

If there is always a state of affairs at least in principle available to the agent at $t$ in which neither the agent’s practical reasoning nor the agent’s behaviour at $t$ is morally wrong, then this definition guarantees that the logic of what is morally required is so-called standard deontic logic (SDL) – in effect, the modal system known as KD.

SDL is controversial, but for our purposes the controversial features of SDL are not important. These controversial features flow from the right-to-left half of this definition (the thesis that whenever your $φ$-ing is necessary for you to avoid moral wrongness, you are morally required to $φ$). What matters for our purposes is the left-to-right half of this definition – the thesis that if you are morally required to $φ$, then it will be impossible for you not to $φ$ without either your behaviour or your reasoning being in some way morally wrong. This left-to-right half of the definition is much less controversial; and in the rest of this discussion, it is only this left-to-right half that I need to assume here.

---

\(^4\) For an argument in favour of this assumption, see John Broome (1999).

\(^5\) I have formulated this definition of moral requirements in this way to allow that moral wrongness can be exemplified in the agent’s practical reasoning – that is, in the way in which the agent forms and revises her intentions or plans for action – as well as in behaviour.
Given what I said in the previous section about the all-things-considered practical ‘ought’, it clearly need not always be the case that whenever one behaves otherwise than as one all-things-considered ought to, one’s behaviour is morally wrong. All things considered, I ought to buy a new pair of shoes; but it surely need not be morally wrong of me not to buy any new shoes.

In what follows, I shall assume a particular conception of moral wrongness – a conception that is inspired by some of what J. S. Mill says in *Utilitarianism* (1871, chapter 5). According to this conception, there is a fundamental connection between what is morally wrong and what is blameworthy. This connection does not imply that every case of moral wrongness is blameworthy. There are certainly cases where an agent’s behaviour is morally wrong, but the agent is blameless because of having some appropriate excuse (such as blameless ignorance or the like). What an excuse of this sort does is to make it the case that even though the agent’s behaviour is in some way morally wrong, the agent is not blameworthy for that behaviour.

However, it is only such excuses that can prevent wrongness from implying blameworthiness. If an agent’s practical reasoning or behaviour at \( t \) is morally wrong, then, unless he or she has an excuse, the agent will be blameworthy for that reasoning or behaviour. So there is still a crucial connection between moral wrongness and blameworthiness.

In Section 1, I interpreted our central question as concerning two propositions, which I numbered (i) and (ii). The second of these two propositions is (ii) – the proposition that whenever we are subject to a non-trivial moral requirement, we have compelling or overriding reasons for conforming to it. If this second proposition is true, then presumably, it is also true that whenever we are subject to a moral requirement, at least some of the reasons that we have for conforming to the requirement ground or explain our being subject to this requirement. These reasons could be called moral reasons. (If we use the term in this way, then even if you are morally required to \( \phi \), there could still be non-moral reasons that count in favour of your \( \phi \)-ing – it is just that such non-moral reasons do not ground or explain your being subject to this moral requirement.)

If such moral reasons exist, it is plausible that there are also other moral reasons that do not ground any moral requirements. First, moral reasons that are themselves overridden by countervailing reasons presumably do not ground any moral requirement: for example, I might have a moral reason to keep my promise to meet you for coffee in the Common Room, but this reason will not ground a moral requirement if it is overridden by the reason that I have to help a badly injured person get to hospital without delay.

Secondly, some moral reasons are merely supererogatory considerations: for example, I have a moral reason to give up my expensive habit of buying opera tickets and to donate all the money to charity instead; but (I hope) this moral reason does not ground any full-blown moral requirement. In cases where a supererogatory option is available, the moral reason in favour of the supererogatory course of action does not “override” the non-moral reasons against that course of action. As we saw in the previous section, to say that your reasons for \( \phi \)-ing are
“overriding” or “compelling” is to say that φ-ing is not merely a good thing to do, but something that you – in the “all-things-considered” sense – ought to do. But even though the supererogatory course of action is not something that one ought not to do, it is surely also not something that one ought to do (in this “all-things-considered” sense). So the moral reason in favour of the supererogatory course of action does not override the reasons in favour of the alternative (even though the reasons in favour of the alternative presumably do not override the reasons in favour of the supererogatory course of action either). It seems clear, then, that these supererogatory moral reasons also do not ground any moral requirements.

These points are important for understanding the second of the two propositions that we are concerned with here, (ii). This proposition implies that the reasons supporting conforming to moral requirements are always overriding or conclusive; it does not imply the much stronger and less plausible proposition that absolutely all moral reasons are overriding.

4. The moral / non-moral distinction

It seems that the question that we are concerned with here would be much less interesting and important if absolutely all reasons for action were moral reasons.

Suppose that all reasons for action were moral reasons. Admittedly, we would still need to explain (i), the first of the two propositions that we are concerned with; that is, we would still need to explain why any moral reasons had the necessary qualities to ground a full-blown moral requirement (given that if you are morally required to φ, then φ-ing is necessary for avoiding wrongness – that is, for avoiding the kind of reasoning or behaviour that, at least in the absence of any excuse, would be blameworthy). But if one is morally required to φ, then it is not surprising that the moral reasons in favour of φ-ing outweigh or override all moral reasons for acting otherwise. So, if all reasons are moral reasons, it is also not surprising that the reasons in favour of conforming to a moral requirement override absolutely all countervailing reasons whatsoever, and so count as overriding or conclusive reasons; that is, it is not surprising that (ii), the second of the two propositions that we are concerned with, is true.

In this way, if there are no non-moral reasons pulling against the moral reasons, our central question would lose much of its apparent importance and urgency. So we need to make it plausible that there really is a distinction to be drawn between moral and non-moral reasons.

Within the framework of the conception of reasons that was articulated in Section 2, the distinction between moral and non-moral reasons corresponds to the distinction between moral and non-moral values. You have a moral reason to take a course of action if and only if that course of action is better than the appropriate benchmark with respect to some moral value – while there is a non-moral reason in favour of a course of action if and only if that course of action is better than the appropriate benchmark with respect to some non-moral value. But what exactly is this distinction between moral and non-moral value?
I shall not commit myself to any specific account of this distinction between moral and non-moral values here. Presumably, a full account of this distinction would have to say something about the connection between moral values and the reactive attitudes, like blame, which were mentioned in the previous section. However, defending any such account would require a thorough investigation, which would take us too far away from our main theme. Here, I shall just canvas a couple of interpretations of this distinction in order to fix ideas.

The first interpretation of the distinction that I shall canvas here is a view that I shall call “Sidgwickian consequentialism”.\(^6\) This view assumes a consequentialist conception of the value of action. That is, this view assumes that the only relevant kind of value that an action can have is a value that is derivative from, and wholly determined by, the value of the action’s total consequence – where the “total consequence” of an action is simply the conjunction of all states of affairs that (i) would obtain if the action were to be performed, and (ii) might not obtain if the action were not performed.

However, Sidgwickian consequentialism differs from many forms of consequentialism in that it does not only focus on the agent-neutral value of consequences. Instead, Sidgwickian consequentialism implies that while some reasons for action are grounded in agent-neutral values, there are also other reasons for action that are grounded in agent-relative values.

Specifically, according to Sidgwickian consequentialism, there are two ways in which reasons for action are grounded in values. First, if the consequences of a course of action that is available to you are good in an agent-neutral way, this fact will ground a reason for you to take that course of action. Secondly, if the consequences of a course of action are good for you, that fact can also ground a reason for you to take that course of action. According to this Sidgwickian view, it is natural to identify the reasons that are grounded in the agent-neutral values with the moral reasons, and to identify the reasons that are grounded in agent-relative values of this sort with reasons of self-interest – where reasons of self-interest are plausibly categorized as non-moral reasons. On this Sidgwickian view, then, it is just a fundamental fact that the values that generate reasons for action include both agent-relative and agent-neutral values, and that in consequence we have both moral and non-moral reasons for action.

The second interpretation of the moral / non-moral distinction that I shall canvas here differs from Sidgwickian consequentialism in two fundamental respects. First, this second interpretation rejects the consequentialist conception of the value of actions.\(^7\) Instead of implying that the value of an action is derivative from and determined by the value of its total consequences, this view permits actions to have an intrinsic value or disvalue of their own – so that it is possible, at least

\(^6\) This view is broadly inspired by Sidgwick (1907).

\(^7\) Although this second interpretation is emphatically non-consequentialist, it is not deontological in the strict sense of advocating principles of rightness that have nothing to do with any notion of moral value or goodness. The value-based conception of reasons for action that I am assuming here makes it hard to reconcile any strictly deontological conception of morality with the belief in the existence of moral reasons.
in principle, for there to be two actions that differ in value from each other even though the overall value of the total consequences of the two actions is exactly the same. (For example, if one of these actions actively causes a bad consequence while the other action merely allows a bad consequence to occur, then the first action may be worse than the second, even if the overall value of their total consequences is the same.) Secondly, according to this interpretation, all reasons for action are grounded in the intrinsic agent-neutral values that are instantiated by the acts themselves – there is no need to appeal to agent-relative values here.

More specifically, according to this interpretation of the moral / non-moral distinction, moral values are all fundamentally instantiated by interpersonal relations. For this reason, I shall call this interpretation the “relational view”. You have a moral reason in favour of a certain course of action if and only if that course of action would put you into a morally good relation to some person or persons; you have a moral reason against a course of action if and only if that course of action would put you into a morally bad relation to some person or persons. For example, an action that involves saving a person’s life will normally put you into a morally good relation to that person, while an action that involves killing a person will normally put you into a morally bad relation to that person.

In general, it seems plausible that the moral value or disvalue of a relation between persons is sensitive to the effect of that relation on the values that ground the other reasons that the persons in question have. Morally good relations between persons often help at least some of those persons to exemplify or promote those values, while morally bad relations between persons typically worsen the position of at least some of those persons to exemplify or promote those values. For example, you have a reason to want to live a long flourishing life – a reason that is grounded in some value, such as the value of well-being or the like. So, if I save your life, the relation between us has helped you to exemplify this value to a greater degree, whereas if I allow you to die prematurely, this relation has, at least by comparison, put you in a worse position to exemplify this value. Proponents of the relational view could say that this difference is part of what explains why I stand in a morally better relation to you if I save your life than if I allow you to die prematurely – and also why I have a moral reason to save your life rather than to allow you to die prematurely.

---

8 This idea is akin to the central idea of Scanlon (1998, 162) that moral reasons are grounded in the “value and appeal” of a certain “relation” that we might stand in to others; it is also akin to the conception of “second-personal” reasons that is advocated by Darwall (2006). However, this view is not committed to any sort of “contractualism” that is built around the idea of principles that could not be reasonably rejected or the like.

9 There seem to be moral reasons that do not arise from the relations between the agent and any person who counts as either a victim or a beneficiary of the agent. For example, there seem to be moral reasons for not despoiling the natural environment, or for saving species or cultural traditions from becoming extinct; and acting against these moral reasons need not involve wronging any individual victims at all. Still, if one acts against these reasons, then one has arguably impaired one’s relation to the whole “moral community”, and it may be this that explains why there is a moral reason (as well as a non-moral reason) not to act in these ways.
As I said, in this section I have simply canvassed a couple of interpretations of the moral / non-moral distinction, for the purpose of fixing ideas about what is at stake in this debate.

In general, on any version of the value-based conception of reasons for action, there will be no great difficulty in explaining why you have moral reasons for action. As long as there are moral values, and some of the courses of action available to you instantiate these moral values to different degrees, this is enough to explain why you have moral reasons for action. Within this framework, there is a moral reason in favour of a course of action if and only if that course of action is better, with respect to one of these moral values, than the relevant benchmark.

So far, however, this framework can only explain why we have some moral reasons for action. So long as there are non-moral reasons as well as moral reasons, it could still be the case, for all that has been said so far, that moral reasons are extraordinarily weak reasons – reasons that sometimes conflict with non-moral reasons for action, and lose out to the non-moral reasons whenever any such conflict arises.

Intuitively, it seems that if moral reasons were such extraordinarily weak reasons, they would be trivial; and similarly, any moral requirements grounded in these moral reasons would be equally trivial. I shall assume that to explain why we are subject to “non-trivial moral requirements” (as I put it in (i) the first of the two propositions that I enumerated at the end of Section 1), we must explain why some moral reasons – specifically, the reasons that ground moral requirements – are not trivial in this way. On the contrary, there are cases in which we can have moral reasons to do something even though we have reasons for being tempted not to do it, and in many of these cases, these reasons are not overridden or outweighed by those countervailing reasons.

To solve our problem, then, we must do much more than simply explain why there are moral reasons. We need to explain (i) why there are non-trivial moral reasons of this sort, which ground correspondingly non-trivial moral requirements, and (ii) why it is the case that whenever we are subject to such a moral requirement, we have an overriding or conclusive reason to act accordingly. We still do not yet understand how that can be the case.

5. The required / supererogatory distinction

At this point, it is illuminating to invoke a point that has perhaps been most clearly formulated by Sarah Stroud (1998).10 As we have seen, there is a crucial difference between (a) moral reasons in general and (b) moral requirements. Some moral reasons are overridden or defeated by other reasons; and these defeated reasons do not ground any moral requirements. Moreover, moral reasons (even undefeated moral reasons) include supererogatory considerations, which also fall short of grounding full-blown moral requirements. So, what distinguishes between those

---

10 Closely related points have also been made by other philosophers, such as Seana Shiffrin (1999).
(undefeated) moral reasons that ground full-blow moral requirements, and those that are mere supererogatory considerations instead? How should this distinction be drawn?

I propose the following account of the distinction: the difference between those actions that are morally required and those that are merely supererogatory consists precisely in the fact that the moral reasons in favour of the former actions are overriding reasons, while the moral reasons in favour of the latter actions are not overriding reasons, but only sufficient reasons for action.

This proposal is supported by the fact that whether we regard the reason for a certain action as grounding a moral requirement or not typically depends on the strength of the reasons (including non-moral reasons) against that action. In other words, when we judge whether a moral reason grounds a moral requirement or is merely supererogatory, we always take account of the other reasons that the agent has. For example, if the cost in terms of the agent’s other reasons for action is exorbitant, we are more likely to classify the moral reason as supererogatory rather than as grounding a moral requirement; whereas if the cost in terms of the agent’s other reasons for action is relatively trivial, then even quite a weak moral reason may ground a full-blown moral requirement. The account of the required / supererogatory distinction that I have just articulated could give a straightforward explanation of this.

A further consideration supporting this account is the fact that it seems intuitively doubtful whether it could be appropriate to blame people for acting in a way in which they had a sufficient reason for acting. If this intuition is correct, then you can only be blameworthy for an action if you had an overriding reason not to perform that action. Given that the notion of moral requirements and moral wrongness has an intimate connection to blameworthiness, this intuition also supports the proposal that I am making in this section.

Clearly, this proposal explains the second of the two propositions that we began with – that is, it explains why it is the case that whenever you are subject to a moral requirement, you have overriding or conclusive reasons for conforming to it.

However, even if this proposal is correct, it only entails that if we are subject to any moral requirements, we have overriding reasons to comply with those requirements. This proposal does not entail that we are subject to any non-trivial moral requirements. It entails that a moral reason for a course of action cannot ground a moral requirement (as opposed to counting as a mere supererogatory consideration) unless it overrides all the reasons against that course of action. But why are there any non-trivial moral reasons that have the power to override all countervailing reasons in this way?

In effect, then, the second of the two propositions that I started with – the proposition that whenever we are subject to a moral requirement, we have overriding reasons to conform to it – turns out not to be the most challenging proposition for us to explain. It is rather the first proposition – the proposition that we are subject to non-trivial moral requirements at all – that is the most challenging problem for us to solve. To solve the problem, we will have to explain why
there are any moral reasons that have the power to override all countervailing non-moral reasons for action.

6. Alternatives to the weighing model?

Some philosophers might be tempted to attack this problem by claiming that non-moral reasons never really conflict with moral requirements at all. The claim that non-moral reasons never conflict with moral requirements could be defended in many ways. For example, some versions of *eudaemonism* seem committed to a claim of this sort. These versions of eudaemonism accept the following two theses: first, there is a reason for you to take a course of action if and only if that course of action contributes to your well-being to a sufficient degree; secondly, no course of action that involves failing to conform to a moral requirement can contribute to your well-being to a sufficient degree.

An alternative version of this approach might concede that in some cases, courses of action that involve contravening a moral requirement may indeed make a significant contribution to the agent’s well-being. Nonetheless, this alternative version of the approach might insist that in these cases, there is no reason for the agent to take any such course of action. In this way, this version of the approach could hold on to the central idea that no genuine non-moral reasons ever conflict with any moral requirement. In some way, the fact about the course of action in question that would normally ground a reason for action is somehow “silenced” by the presence of the moral requirement; in other words, the moral requirement “disables” that fact from generating any genuine reason for action.

For our purposes, these differences between the various ways of denying that there are ever any non-moral reasons against conforming to a moral requirement do not matter. All that matters is that these approaches are committed to the thesis that none of the apparent benefits provided by the agent’s failing to conform to a moral requirement can ever ground a genuine reason against conforming to that requirement.

Given some extremely plausible assumptions, this thesis implies that there is a radical discontinuity among the apparent benefits in question. These apparent benefits normally ground a reason for acting; but however great these benefits may be, as soon as there is moral requirement not to act in the way that secures those apparent benefits, they suddenly cease to ground any reason for action at all.

---

11 For some discussion of the sort of eudaemonism that seems to have been assumed by Plato, see Wedgwood (2009b).

12 Compare T. H. Irwin’s (1994) interpretation of Plato as claiming that justice is a “dominant component of happiness”.

13 Compare John McDowell (1979) and Jonathan Dancy (2005, 41).
Such discontinuity claims have a grandiose sound to them, but they typically reveal themselves on closer inspection to face severe problems. The main reason for this is that everything that matters in life comes in degrees. The factors on which both reasons and values – including both moral and non-moral reasons and values – are grounded seem to be capable of varying in arbitrarily small increments from one case to another. For example, an act can be morally bad because it causes harm, but causing harm is a factor that can vary by tiny increments. For example, even if one harm is greater than a second, the degree to which the first harm is greater than the second may be extremely small. Indeed it is not clear that there is any limit to how small the difference between a greater and a lesser harm can be; harms seem to be capable of varying by arbitrarily small increments.

We can explore the problems with these discontinuity claims by imagining the following spectrum of cases, $C_1, \ldots, C_n$. In each case $C_i$, there are two options available to you, $A_i$ and $B_i$. There is at least one case $C_k$ in which there is indisputably both a moral reason in favour of $A_k$ and a non-moral reason in favour of $B_k$; and throughout this spectrum, the factors on which these reasons depend in $C_k$ vary only very slightly between each case and its successor. But intuitively in the first case $C_1$ the moral reasons in favour of $A_1$ seem overriding, while in the last case $C_n$, the non-moral reason in favour of $B_n$ seems clearly not to be overridden by any countervailing moral reason.$^{14}$

To give a more concrete illustration of such a spectrum of cases, assume that in the case $C_k$ where there is clearly both a moral reason and a non-moral reason, the moral reason in favour of $A_k$ is grounded in the degree to which $A_k$’s total consequences is good for the world, while the non-moral reason in favour of $B_k$ is grounded in the degree to which $B_k$’s total consequences are good for the agent. For example, suppose that in the first case $C_1$, option $A_1$ involves your saving millions of others from an agonizing death but not saving yourself from a paper cut, while option $B_1$ involves your saving yourself from a paper cut but not saving millions of others from an agonizing death. In each subsequent case $C_i$, the harm that others will suffer if you do $B_i$ is slightly less than in the preceding case, while the cost that you will suffer if you do $A_i$ is slightly greater than in the preceding case. Finally, in the last case $C_n$, option $A_n$ involves sacrificing your own life to save the life of a distant stranger, while option $B_n$ involves saving yourself but not saving that the life of that distant stranger.

The following claims about these cases seem plausible:

a. In case $C_1$ you are morally required to choose option $A_1$;
b. In case $C_n$ you are not morally required to choose option $A_n$;
c. There is a case $C_j$ on this spectrum such that $1 \leq j < n$, where $C_j$ is the last case in which you are morally required to choose the $A$-option;

---

$^{14}$ This sort of spectrum of cases is inspired by those that are deployed by Williamson (2000, chapter 4).
d. There is at least one case $C_k$ such that $j < k \leq n$, where option $A_k$ is supererogatory but not morally required.

The views that we are considering in this section must all posit a radical discontinuity on this spectrum of cases immediately after case $C_j$ – in spite of the apparent continuity of this spectrum of cases. But if there really is a radical discontinuity on this spectrum, then in the crucial respects this is not really a continuous spectrum of cases at all. $C_j$ is a special point on the spectrum, since immediately after $C_j$ you cease to be morally required to help others rather than yourself; and after $C_j$, the fact that one of the available courses of action will spare you physical harm suddenly jumps up from not grounding any sort of reason for action at all, to grounding a very significant reason. It seems impossible to explain why these factors would behave in such a radically different way in the two cases on either side of this special point, given how extremely similar to each other these two cases seem to be. So, views of this sort – which as we have seen include both the eudaemonist view and the silencing view – seem implausible.

So, it seems, there must always have been a non-moral reason in favour of the $B$-option in all of the cases on spectrum, even in cases in which you were morally required to choose the $A$-option. The non-moral reason in favour of the $B$-option was really present in all these cases, conflicting with the moral reason in favour of the $A$-option; it is simply that in those cases this non-moral reason was not strong enough, and the moral reason was strong enough, to ensure that the non-moral reason in favour of the $B$-option was overridden by the moral reason in favour of the $A$-option.

It seems then that the best way to explain these phenomena is by invoking a simple weighing model. According to this model, in each case $C_i$, the totality of reasons in favour of $A_i$ and the totality of reasons in favour of $B_i$ both have a weight – where these weights can be at least partially ordered, and out of a given set of alternative acts $\{A, B \ldots\}$, there is overriding reason in favour of $A$ if and only if the totality of reasons in favour of $A$ is weightier than the totality in favour of each alternative.

The weighing model can easily explain everything that it is plausible to say about this spectrum of cases, given the assumption that – at least in these cases – the weight of the reasons in favour of each act is an increasing function of the harm that that act prevents. In $C_1$, the reasons in favour of $A_1$ are clearly weightier than the reasons in favour of $B_1$. Then, as we go along the spectrum of cases, the reasons in favour of the $A$-option get gradually less weighty, while the reasons in favour of the $B$-option get gradually weightier, until we pass case $C_j$, after which point the reasons in favour of the $A$-option cease to be weightier than the reasons in favour of the $B$-option (which is not to say that after that point, the reasons in favour of the $B$-option become weightier than the reasons in favour of the $A$-option – indeed, in case $C_k$, where option $A_k$ is supererogatory, neither the totality of reasons in favour of $A_k$ nor the totality of reasons in favour of $B_k$ is weightier than the other).
Some philosophers may be tempted to rescue the “silencing” approach by appealing to Joseph Raz’s notion of an “exclusionary reason”.\textsuperscript{15} In the cases in which one is morally required to help others rather than oneself, perhaps there is – in addition to a moral reason to help the others – an exclusionary reason against even considering any countervailing reasons. However, it is doubtful whether there is always such an exclusionary reason in all these cases. If the case is very close to the line between the cases where helping others is morally required and the cases where helping others is morally supererogatory, then there may not be a reason against your considering the reason to help yourself rather than the others.

Moreover, even if there always is such an exclusionary reason, it will often be such a weak reason that it is itself overridden by countervailing reasons or other factors. So this appeal to exclusionary reasons seems to presuppose that the exclusionary reason is itself a compelling or overriding reason. But why should this exclusionary reason count as compelling or overriding in this way? Clearly this is a question of exactly the same kind as the fundamental problem that we are seeking to address. So, even if there is always an exclusionary reason against even considering reasons for courses of action that involve contravening a moral requirement, it seems that the appeal to exclusionary reasons does not solve the fundamental problem that concerns us.

It seems, then, that we cannot solve our problem by denying that non-moral reasons ever conflict with moral requirements. This approach fails to recognize that both moral reasons and non-moral reasons come in degrees, and that some moral reasons are stronger or weightier than others. There are, at least sometimes, genuine non-moral reasons in favour of courses of action that involve failing to conform to a moral requirement.

Instead, then, I propose, when moral reasons override non-moral reasons, it is simply because they outweigh them. But why do the moral reasons outweigh the countervailing non-moral reasons in cases of this kind?

7. Moral values as big values

Giving a complete theory of what determines the weight of reasons would be far too large a task for me to attempt here. So instead, in this final section, I shall enumerate some of the factors that seem to be involved in determining how much weight a given reason has. Once these factors are in view, they may help to make it seem less surprising and puzzling that moral reasons often outweigh countervailing non-moral reasons.

The kind of “weight” that I have just been talking about attaches to the totality of the reasons in favour of each of the available acts; it does not attach to any of the individual reasons belonging to this totality. Nonetheless, it may be that in at least some cases, this weight can itself be

\textsuperscript{15} See Raz (1999b, 35-48).
regarded simply as the *sum* of the weights or strengths of all the various individual reasons in favour of the act in question.

Some philosophers will be tempted to think that this “additive” version of the weighing model is hopelessly crude; these philosophers may insist that the weight of the totality of reasons in favour of an act may in some cases be determined by *complex interactions* between the various individual reasons in favour of the act and in favour of its alternatives, and so cannot be identified with the sum of the weights of all the individual reasons. However, we could always say that the effect of these “complex interactions” is simply to generate new reasons in favour of (or against) some of the available acts, in which case the presence of these new reasons will obviously affect the sum of the weights of all the reasons. So it is not clear that it is such a crude mistake to assume an additive version of the weighing model.

According to this additive version of the weighing model, the crucial question will be what determines the weight of the individual reasons in favour of each of the available acts. As I explained in Section 2 above, I am working here within the framework of a value-based conception of reasons for action. According to this value-based conception, there is a reason in favour of an act if and only if that act is better (with respect to one of the relevant values) than the relevant benchmark – which implies that this act is better than at least some available alternatives.

Now, there is one factor that seems normally to make a crucial difference to the weight of a reason: at least other things equal, the greater the degree to which an action is better (with respect to the relevant value) than the available alternatives, the weightier the reason for that action is. If there is only a small difference between the available acts with respect to a suitable value, then (other things equal) this value will ground only a relatively weak reason in favour of the acts that are better than the relevant benchmark with respect to that value. If there is a huge difference between these acts with respect to the value, then (other things equal) this value will ground a much stronger reason in favour of the acts that are better than the benchmark with respect to that value.

As I described this factor, however, it only affects the relative weights of different reasons that are grounded in the same value. If the reason in favour of an act $A$ is grounded in the fact that $A$ is better than the relevant benchmark with respect to a certain value, then other things equal, this reason would have been stronger if the degree to which $A$ is better than the benchmark with respect to that value had been greater than it actually is.

One might attempt to extend this point to reasons that are grounded in different values. Even if a reason in favour of $A$ and a reason in favour of $B$ are grounded in two different values $V_A$ and $V_B$, could it sometimes still be the case that the degree to which $A$ is better than the benchmark in terms of $V_A$ is greater than the degree to which $B$ is better than the benchmark in terms of $V_B$? If
so, then perhaps it might be true here too that other things equal, the reason in favour of \( A \) grounded in \( V_A \) is weightier than the reason in favour of \( B \) that is grounded in \( V_B \).

However, there are grounds for doubting whether this is a promising approach.\(^{16}\) It seems that it does not always make sense to compare the difference between two items with respect one value with the difference between a pair of items with respect to another value. Is the difference in musical value between Beethoven’s *Missa Solemnis* and Ravel’s *Bolero* greater or lesser than the difference in hedonic value between the typical migraine that was suffered by Virginia Woolf and the agony of those who were crucified to death by the ancient Romans? It is not clear that such comparisons make any sense at all; even when all of the items being compared are acts, it is still not obvious that such comparisons can really be made. (Of course, we could make such comparisons if by saying that the difference between \( A \) and the benchmark in terms of one value \( V_A \) is “greater than” the difference between \( B \) and the benchmark in terms of a second value \( V_B \), we meant no more than that the reason in favour of \( A \) grounded in \( V_A \) is weightier than the reason in favour of \( B \) that is grounded in \( V_B \); but in that case, this comparison obviously could not explain why the first reason is weightier than the second.)

Still, there may be another factor that applies to reasons that are grounded in different values, which explains why reasons that are grounded in one value \( V_1 \) are often weightier than reasons grounded in a second value \( V_2 \). This factor is present most clearly in cases where the fact about the first value \( V_1 \) that grounds the first reason in a sense “subsumes” the fact about the second value \( V_2 \) that grounds the second reason.

We can illustrate this point by considering the Sidgwickian form of utilitarianism. According to this form of utilitarian, a *reason of self-interest* in favour of an act \( A \) is ultimately grounded in the degrees to which the state of affairs of the agent’s doing \( A \) is better for the agent than each of the relevant alternatives. By contrast, a *moral reason* in favour of an act \( B \) is grounded in the way in which the state of affairs of the agent’s doing \( B \) compares to each of the relevant alternative acts in terms of *sum* of the degrees to which the act is better or worse than each of the relevant alternatives for every individual person who will exist if the act is performed. In effect, as we might put it, the evaluative fact that grounds the moral reason is *partially constituted* by the evaluative fact that grounds the reason of self-interest.

For example, suppose that I set off a bomb, killing myself and hundreds of other people. According to the Sidgwickian form of utilitarianism, the reason of self-interest against this act is grounded in a fact that is constituted by the effects that this act has on my well-being, while the moral reason against this act is grounded in a fact that is constituted by the effects that this act will have on my well-being *and* on the well-being of everyone else as well.

The similar point will hold for many other consequentialist views besides utilitarianism. The degree to which the consequences of the available acts differ in how good they are for the

---

\(^{16}\) I am indebted to an anonymous referee for convincing me that this approach is not really promising.
individual agent is constituted solely by the effects of those consequences on the life of the agent alone, while the degree to which they differ in the relevant agent-neutral intrinsic value is constituted by their effects on the world as a whole (including the life of the agent). The point seems to lie behind an observation that Aristotle makes in the *Nicomachean Ethics* (1094b5-6):

> For even though the human good is the same for an individual and for a community, the good of the community is manifestly a greater and more perfect good both to attain and to preserve. To secure the good even for one person is worthwhile, but it is a finer and more divine thing to secure the good for a nation or for communities.

A similar point may well be true on the relational view as well. The relational view rejects the consequentialist conception of the value of courses of action, but it still interprets the moral value of an act as arising from the effect of the relationship that that act creates between two or more agents on the values that ground the reasons that those agents have. In effect, the moral value of this relationship between agents is partially constituted by the facts about values that ground those other reasons that those agents have.

On the relational view, then, the facts about moral value that ground the moral reasons that each individual has in this sense “subsume” the facts about non-moral values that ground the reasons that are possessed by several individuals. For example, according to this view, the moral reason that I have for saving your life if I can do so at negligible cost to myself is grounded in the fact that I put myself into a morally better relation to you if I save your life than if I fail to do so; and that fact is partially constituted by whatever fact about values (such as the value of your well-being) grounds the reason that you have for wanting to go on living.

If it is true that moral values “subsume” other values in this way, then, as we might put it, moral values are “big values” – values that put together the facts about other values to make a bigger more encompassing evaluative fact.17 According to the tentative proposal that I wish to make here, the fact that moral values are “big values” in this way is at least part of what explains why moral reasons are often so weighty that they outweigh all countervailing non-moral reasons.

At the same time, it seems that there must be other factors affecting the strength of a reason besides how big the difference is between acts (in terms of a single value), and how “big” the relevant value is. As we have seen in cases of supererogation, moral values do not always outweigh countervailing non-moral reasons. My reason to promote my own well-being seems much weightier than the reason that I have to take equally burdensome steps to promote the well-being of a stranger who is wholly unknown to me, and the reason for this seems to have to do with factors of a completely different kind. For example, this factor might be something like the *motivational centrality* of the value of one’s own well-being in the mental life of normal human

---

17 For the contrast between “big values” and “small values”, see Raz (1999b, 30).
beings; or perhaps it has to do with the proximity of the state of affairs that fundamentally exemplifies this value to the practical situation of the agent in question.

At all events, the central proposal that I wish to make here is that the weight of moral reasons is explained, at least in part, by the fact that moral values are big values in this way. However, this proposal cannot quite be the whole of the story. For this fact about the moral values to give reasons for us, we need to have the right sort of opportunities: an appropriate array of alternative acts must actually be available to us, in the many situations in which the moral reasons emerge as overriding or compelling reasons for action. This is because it is only when the acts that are available to us differ sufficiently greatly in their moral value that we have the particularly weighty moral reasons that I have been describing here.

What explains why the opportunities available to us are so often of this sort? This is ultimately an empirical question, about the explanation of a contingent fact about human beings. Still the following speculation seems plausible: the explanation has to do with the radically social nature of human beings – that is, with the fact that the evaluatively significant aspects of our lives consist to such a great extent in our interactions with other persons. If there were intelligent agents who were less deeply social than we are, they might rarely be in situations in which the available acts differed greatly in moral value, while they might more often be in situations in which the available acts differ greatly with respect to non-moral values; and it seems that these non-social agents would be subject to less demanding moral requirements than we are.

At all events, an approach along these lines seems to be the most promising way for a value-based conception of reasons for action to explain why it is that we human beings are subject to non-trivial moral requirements, which correspond to overriding reasons for action. The acts available to us often differ significantly in their moral values; and moral values are, as I put it, particularly big values. This may be what explains why we so often moral reasons for action that are weighty enough to outweigh all the reasons that we have for acting otherwise.18

References


18 An earlier version of this paper was presented at the University of Arizona’s Normative Ethics Workshop, and at the Central Division Meeting of the APA in Chicago. I am grateful to those audiences, and also to Jonathan Dancy, Michael Otsuka, Joseph Raz, and an anonymous referee for this volume, for helpful comments.


