University of Reading

# *Should philosophical reflection on ethics do without moral concepts?*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open access

Hooker, B. (2024) Should philosophical reflection on ethics do without moral concepts? Ethical Theory and Moral Practice, 27. pp. 651-665. ISSN 1572-8447 doi: 10.1007/s10677-022-10303-x Available at https://centaur.reading.ac.uk/106612/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1007/s10677-022-10303-x

Publisher: Springer

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

Ethical Theory and Moral Practice
https://doi.org/10.1007/s10677-022-10303-x

# Should Philosophical Reflection on Ethics Do Without Moral Concepts?

**Brad Hooker**[1]

Accepted: 23 May 2022
© The Author(s) 2022

**Abstract**
Roger Crisp, in his book *Reasons and Goodness*, argues in favour of de-moralizing our philosophical reflection on ethics. This paper begins by explaining what 'de-moralizing' means. Then the paper assesses Crisp's argument for de-moralizing and puts forward arguments against de-moralizing.

**Keywords** Moral Concepts · Benevolence · Coordination rules · Justice · Fairness · Rights · Loyalty · Desert

Roger Crisp's *Reasons and Goodness* (2006) contains innumerable insights and very many compelling arguments.[1] However, the book's argument for de-moralizing our philosophical reflection about ethics seems to me unpersuasive. This paper explains why.

## 1 Ethics and Morality

The suggestion that philosophical reflection on ethics could be de-moralized presupposes that there are ethical concepts that are not moral ones as well as ethical concepts that are moral ones. The concepts Crisp classifies as moral are the concepts of *moral reason*, *moral prohibition*, *moral requirement*, *moral obligation*, *moral duty*, *moral prerogative*, *moral rightness*, *moral wrongness*, *virtue*, *kindness*, *cruelty*, and *justice* (pp. 1, 13, 25). One of the central contentions in Crisp's book is that our theorizing about ethics will go better if we eschew appeal to, or reliance on, moral concepts.

[1] Henceforth, references to this book will be in my text by page number or chapter number.

✉ Prof Brad Hooker
b.w.hooker@reading.ac.uk

[1] University of Reading, Reading, UK

Published online: 16 June 2022

There is not a sharp line between ethical theorizing and everyday ethical thought (Hooker 2018). Nevertheless, it is one thing to contend that philosophical reflection on ethics should try to build up from a pared down set of concepts, and it is another thing to contend that people's everyday ethical thought about what to do and how to react should be purged of these same concepts. For the sake of argument, suppose *everyday ethical thought* works far better because it utilizes moral concepts than it would if it tried do without moral concepts. That supposition does not entail that ethical *theorizing* must draw on, rather than sidestep, moral concepts. Crisp himself writes, 'avoiding the moral concepts in philosophical reflection can sit quite happily with making the strongest moral condemnations in the world outside philosophical reflection.' (p. 27).

Philosophical reflection on ethics can take many forms, one of which is trying to ascertain the ultimate, non-derivative, principles about reasons. One such principle is that all agents have *non-derivative* reasons to make their lives go better overall for themselves, or, in other words, to promote their own individual well-being (p. 73). The idea that there is a non-derivative reason to promote one's own well-being is a normative idea, but it does not involve any moral concept, unless the concept of well-being is itself moralized. *Well-being* is a moralized concept if but only if at least some of what constitutes well-being depends on moral distinctions. Many people think that morally innocent pleasures and morally innocent achievements add to an agent's own well-being but sadistic pleasures and evil achievements don't. These people have a moralized concept of well-being.

Another intuitively compelling principle, according to Crisp, is that all agents have non-derivative reasons to promote impartial well-being, where the well-being of each, including oneself, counts equally (pp. 141–42). This principle can take two forms, one where *well-being* is not a moralized concept and one where this concept is moralized. If the principle is that all agents have non-derivative reasons to promote impartial well-being on the assumption that well-being is not moralized, then Crisp would call the principle an ethical one but not a moral one, since the principle involves no moral concepts whatsoever. However, many people hold that a particularly obvious *moral* reason is the altruistic reason to do good *for others*. Thus, many people will think that a principle about promoting impartial good is a moral principle. But, for the sake of argument, let's grant to Crisp that a de-moralized conception of ethics is one that (a) begins with dualistic reasons to promote both one's own well-being and impartial well-being, with the stipulation that well-being does not depend on moral distinctions, and then (b) without further bringing in or drawing on any moral concepts, develops a full-blown theory of what to do or not do, what intentions to have or not have, what character traits and reactive attitudes to have or not have, etc.

## 2 Crisp's Arguments from Moral Disagreement and the Distinction Between Derivative and Non-Derivative Reasons

If one is trying to develop a full-blown ethical theory having started from as small a set of concepts as possible, the resulting theory had better not affirm in one place what it denies in another. Even a theory that manages to be fully internally consistent will strike us as unconvincing if it makes claims or has implications that seem to us, after careful reflection, implausible. To be sure, our confidence in various moral ideas might well fade when we see that these ideas are incompatible with some impressive ethical theory, especially if the

theory gives convincing explanations of why these ideas are misguided. But we are hardly likely to accept as justified an ethical theory that conflicts with normative beliefs that seem to us more intuitively compelling than the theory itself seems.

So, will *starting* with the concepts of reasons to promote the well-being of oneself and others and then *going on* to build an ethical theory that eschews all moral concepts yield an ethical theory with sufficiently plausible implications? I agree with Crisp that our most confident normative convictions include the conviction that there are egoistic reasons to promote one's own well-being and the conviction that there are reasons to promote the well-being of others. However, I think that the most confident normative convictions that others share with me also include some convictions containing moral concepts.

There are some very abstract moral ideas that are very difficult not to accept. An example is that everyone matters morally. Admittedly, this idea is too abstract to take us far in deciding which systematic account of ethics is most plausible. However, some abstract convictions, including the example I just offered, have played immensely important roles in practical ethics.

Some other moral propositions are much less abstract and yet seem very difficult not to accept. I have in mind the following:

- There is a *general moral obligation to do good for others.*
- There are *moral prohibitions on certain kinds of act*, such as physically harming the innocent, stealing or destroying others' property, breaking promises, telling lies, treating people unfairly, punishing the innocent, and threatening to do any of these things.
- There are *special moral obligations* to those with whom one has certain special connections, i.e., obligations of loyalty, gratitude, and reparation.
- There are moral *prerogatives* to give one's own good either somewhat *greater* weight (Scheffler 1982, ch. 2), or somewhat *less* weight (Slote 1985, ch. 1), than the equal good of others when one is deciding how to allocate one's own resources.

The intuitive grip of these propositions is enormous, once we construe them as being about obligations, prohibitions, and prerogatives that have limits and that are *pro tanto* rather than necessarily decisive (Hurka 2014, pp. 150–58, 180–81). If we understand the above moral propositions as being about obligations, prohibitions, and prerogatives of limited scope and merely *pro tanto* force, I cannot think of anyone who rejects *all* the moral propositions listed above unless the rejection is driven by a wholescale rejection of morality. Admittedly, each of the above moral propositions has its opponents, even among those who do not reject morality wholescale. Nevertheless, these propositions seem to me to be contents of moral convictions in which most people have great confidence.

My focus here is on beliefs in which not only I but also (I think) most other people have great confidence. The rationale for this focus does not derive from the illusion that consensus guarantees truth. The rationale for my focus instead comes from the conjunction of two assumptions. First, reasoning should typically begin *from* propositions in which one has great confidence. Thus, my own reasoning in ethics should start from the ethical beliefs in which I have the most confidence. Second, in so far as I am putting forward arguments to others that might persuade them, I should put forward arguments with premises that these people accept. If I advance an argument with premises that we all accept, then this argument stands some chance of rationally convincing all of us of its conclusion.

Now, anyone who thinks that we are extremely confident in accepting the obligations, prohibitions, and prerogatives that I listed above will presumably agree that our assessment of any proposed ethical theory must ask whether that theory manages to endorse these obligations, prohibitions, and prerogatives. An ethical theory might endorse them either by (a) taking these obligations, prohibitions, and prerogatives to be the *pluralistic* normative foundation from which the rest of morality can be derived or (b) taking these obligations, prohibitions, and prerogatives to be *derived* from some deeper normative foundation. Which of these two structures of endorsement is more convincing is a further question.

Following many writers, I will refer to the above obligations, prohibitions, and prerogatives, with their limits and pro tanto status, as common-sense morality. Crisp does not have sufficient confidence in common-sense morality to make endorsement of it by a proposed ethical theory a necessary condition of rational acceptance of that theory. Indeed, the motivation for de-moralizing ethics seems to be largely the idea that our theorizing in ethics should not be impeded by a constraint to develop a theory that coheres with common-sense morality.

One of Crisp's reasons for rejecting endorsement of common-sense morality as an epistemic test of proposed ethical theories is that there have been and remain large moral disagreements between people (p. 7). He admits that complete consensus on a proposition's truth is too much to require since no principle in ethics attracts complete consensus (p. 92). But he thinks that, concerning the principles collectively referred to as common-sense morality, the level of disagreement is too great.

There are ethical convictions in which I have great confidence although I know plenty of people do not share these convictions. Is it reasonable of me to have moral convictions I know others don't share? The answer depends on whether I have undefeated (albeit defeasible) arguments or evidence for these convictions.[2] If the answer is yes, then my confidence is reasonable even if others do not share my convictions.

Crisp has another argument against having sufficient confidence in common-sense morality:

> [A]bstract rational intuition will quickly validate many common-sense moral rules. This should not itself be taken to count in favour of such intuition, however, since the intuition is itself insight into the reasons behind those rules. The rules of common-sense morality are not, I shall suggest, in themselves statements of ultimate reasons and hence have no grounding or justificatory weight. (p. 7)[3]

Presumably, Crisp means that we see that the reasons behind the rules of common-sense morality are really merely reasons to promote well-being generally. However, lots of theorists oppose the idea that all the reasons behind the rules of common-sense morality are united in being reasons to promote well-being. Kantians, contractualists, and most virtue ethicists ground the rules of common-sense morality not in the promotion of well-being but in other ideas. Deontological pluralists reject the idea that any one single reason lies behind all the various rules of common-sense morality.

[2] This evidence can include the independent credibility of the content of the conviction. In other words, evidence can include intuitions, conceived of as at least fairly firm intellectual seemings that are not inferred from theories or arguments (Stratton-Lake 2016).

[3] Compare the middle sentence of this passage with Parfit 2017, p. 433.

Even if those kinds of theorists disagree with Crisp that rational intuition reveals that the rules of common-sense morality are justified solely by, and hence completely derivative from, their role in promoting well-being, let us grant that Crisp is correct about that and then note what follows. What follows is only that the rules of common-sense morality are not the ultimate ground, or justificatory foundation, of morality, i.e., that these rules are not morally rock bottom. What does not follow, however, is that the rules of common-sense morality are ethical convictions in which we have less confidence than we do in some more fundamental conviction from which the rules of common-sense morality can be derived. Indeed, if we have more confidence in the rules of common-sense morality than we do in any proposed more abstract moral principle offered as a foundation for normative ethics, then of course we should test the proposed more abstract moral principles by whether they cohere with (by endorsing) the rules of common-sense morality.

## 3 Crisp's argument from parsimony

Crisp has another argument against appeal to moral convictions in our ethical theorising. This argument starts by pointing to:

> *Positive Morality.* A set of cognitive and conative states, including beliefs, desires, and feelings, which leads its possessors among other things to (*a*) avoid certain actions as wrong (that is, morally forbidden) and hence to be avoided, (*b*) feel guilt and/or shame as a result of performing such actions, and (*c*) blame others who perform such actions. (p. 9)

We might say that positive morality consists of shared patterns of *thinking* in terms of wrongness, shared patterns of *feeling* including guilt, shame, and blame, and shared patterns of *acting* including avoiding actions as wrong.

Crisp points out that positive morality is functionally, structurally, and genetically analogous to law. Positive morality and law are functionally analogous in being mechanisms of social control, with law being enforced by police, courts, and punishments, and positive morality being enforced by feelings of guilt, blame, shame, indignation, or resentment. Positive morality and law are structurally analogous in that they both forbid certain kinds of action and create opportunities for other actions (for example the opportunity to create new obligations by signing contracts or making promises). Positive morality and law are genetically analogous in evolving out of negative emotions such as fear of others and anger, emotions which are transformed and channelled so as be mechanisms of social control.

Having highlighted the ways in which law and morality are analogous, Crisp points out that they differ in their relations to non-derivative reasons for action (pp. 14–16). The law does not provide non-derivative reasons for action. One typically has *self-interested* non-derivative reasons for complying with the law, and there might also be *moral* reasons for complying with the law, based on ideas of reciprocity or consent or something else. Thus, the law can be seen as providing derivative reasons derived from non-derivative self-interested reasons or from non-derivative moral reasons. In contrast, positive morality cannot satisfactorily fulfil its functional role as a mechanism of social control unless people believe moral reasons are non-derivative. 'Guilt and moral shame usually involve the thought that

the wrongness of what the agent did itself constituted an ultimate [non-derivative] reason not to act in the way she did. Expressions of blame … are most likely to succeed when the blamed agent feels that the wrongness of what she has done was itself an ultimate reason.' (pp. 15–16).

To illustrate, imagine that Joanna could steal something valuable without being caught. If Joanna thinks the moral prohibition on stealing gives her a reason not to steal *only* if and when stealing is against her self-interest, then she is more likely to steal. If Joanna believes that the moral prohibition on stealing gives her a reason not to steal even when stealing is in her self-interest, she is less likely to steal. This illustrates why positive morality will perform its social control function more effectively if subscribers to this morality believe that it imposes on them reasons that are independent of whether the action would serve their self-interests. Much the same could be said about independence from whether the action would satisfy the agent's preferences.

It is one thing to hold that those who subscribe to positive morality need to believe that moral reasons are ultimate, or categorical, rather than derived from self-interest or personal preference. It is another thing to hold that these people need to believe that the moral reasons imposed by positive morality's obligations, prohibitions, and prerogatives are ultimate in the sense that they are *morally rock bottom*, that is, not derived from some deeper moral principle. If asked why Joanna's stealing would be wrong, for example, would people who subscribe to positive morality answer, 'It just is—there is no deeper explanation and no further reason'? I think most people who subscribe to positive morality would *not* say that. And the reason most people would not say that is that they are often unsure which moral principles are foundational rather than derived, and so unsure whether the prohibition on stealing is morally rock bottom or derived.

I agree with Crisp that positive morality's function is social control. It performs this function more effectively if people believe that moral reasons are not dependent on the agent's self-interested advantage or preference satisfaction. Thus, I accept the inference, 'So it is only to be expected that we who have been brought up to accept a positive morality should tend to believe that morality itself provides reasons.' (p. 16) What I don't accept is Crisp's next step: 'it seems somewhat unparsimonious to postulate genuine properties of wrongness, understood as intrinsically reason-giving, when their attribution by human beings can be fully explained without reference to them.' (p. 16). That benefits flow from people's having some belief does not provide evidence that the belief is true, but nor is it any evidence that the belief is false.

Crisp extends his argument against positive morality to what he calls 'ideal morality', i.e., 'that set of principles, independent of human judgement, which state the truth about wrongness and rightness as ultimate reasons for action' (pp. 15–16). I agree that *positive* morality's existence—as compared with not having mechanisms of social control other than law, etiquette, and custom—brings immensely important benefits. And *ideal* morality cannot be so ideal that it overlooks the real world's need for morality as a mechanism of social control. Both positive morality and ideal morality thus have tight connections to the benefits of social control.

When the beliefs included in a positive morality are ones we share, we think that these beliefs are (as far as we can tell) extensionally equivalent to the propositions of ideal or true morality. If our confidence in common-sense morality is high enough, we might suspect that common-sense morality will be the content of ideal morality. If common-sense morality is

subscribed to by enough other people, then common-sense morality will of course also be the content of positive morality. In contrast, to the extent that positive morality contains elements we reject, then our explanation of why those elements are believed by other people, if we have an explanation, will *not* point to truth makers. And we will *not* be holding that the mistaken elements of positive morality generate ultimate reasons for action (though the need to get along with people committed to the mistaken elements might generate derivative reasons to comply with those elements).

Let us focus now on ideal, or true morality. Is *it*s postulation of ultimate reasons an unparsimonious commitment? Crisp's answer is yes. He rhetorically asks, 'What extra reason against, say, killing in certain circumstances is provided by that killing's being forbidden by some set of special principles?' (p. 16) Suppose *A* could kill *B*, against *B*'s will. The prospective losses in terms of *B*'s autonomy and well-being both impose on *A* reasons not to kill *B* and make *A*'s killing *B* wrong. Why believe that, apart from the reasons constituted by *B*'s prospective losses, the wrongness of killing provides an additional reason on *A* not to kill?

However, suppose *A* is being lethally threatened by *B*. Plausibly, there are still reasons for *A* not to kill *B*, but now these reasons not to kill are opposed by a reason to kill, namely that killing is necessary for self-protection. Principles about wrongness can help adjudicate conflicts between reasons. And when conflicts of a certain kind are common enough, there might be a principle addressing conflicts of this kind, for example the principle allowing killing when necessary for self-defence.

Those who are sceptical about the explanatory value of appealing to wrongness might say that the explanation in the case I've described needs to refer merely to one reason (of self-defence against illicit threat) defeating another (against killing someone). I agree that, with respect to many situations, we can give a fully adequate normative account of what is wrong by saying that one reason or set of reasons outweighed, or in some other way defeated, another reason or set of reasons. If holding that some act would be wrong is holding nothing more than that the balance of reasons in the situation comes out against doing this act, can we not dispense with wrongness and just focus on the balance of reasons?

In order to assess the balance-of-reasons approach, let us consider a possible situation in which you have worked yourself to near exhaustion to help those in need. Suppose that you've done far more than anyone else to help. Suppose also that there are more people whom you could help. Finally, let us suppose that, although you are very tired and hungry, your helping the next such person would benefit her a bit more than your immediately going home to rest, rather than helping her, would benefit you. Presumably, in this situation, the benefit you could give the next person generates a reason to help her, but the benefit to you of going home instead is a reason for you not to do so.

Now if the strengths or weights of reasons that are grounded in prospective benefits are determined *solely* by the sizes of the prospective benefits, such that greater prospective benefits always generate stronger reasons, then your reason to benefit her is stronger than your reason to benefit yourself, since the benefit to her would be a bit larger. However, unless we are act-utilitarians, we think that it would *not* be wrong of you not to help her, since you've already done more than morality can reasonably require of you. Helping her in this case would be supererogatory rather than required (McElwee 2010, p. 315).

Here I might be accused of pointing to a case as an apparent counter-example to Crisp's balance-of-reasons account, when in fact his view would not have the implication I am cit-

ing as counter-intuitive. Crisp holds that 'there is a limit on an agent's reason to sacrifice herself for others that operates when the agent is already below some threshold level of well-being, or the sacrifice called for will take the agent below that, assessed either at-a-time or globally' (p. 141). What he means here by sacrifice assessed globally is the aggregate of sacrifices in iterated cases.

I agree that there are limits on the amount of self-sacrifice that is required for the sake of the impartial good. However, upon careful consideration, a limit specified in terms of reducing one's own good below some level of sufficiency seems to me implausible (Hooker 2009, Sect. 2). *A fortiori*, I find too extreme the idea that a prospective good that would come from helping someone else does not give the agent *any* reason to help her when the agent is already below some threshold for sufficient well-being. But maybe Crisp did not have that extreme idea in mind. Perhaps, what he had in mind is the less extreme idea that, even if there is *some* reason to help the other person in such a case, this reason would be insufficient to outweigh the powerful self-interested reason not to help. If this is what Crisp had in mind, then his balance-of-reasons account could endorse the agent's refusing to make the sacrifice in such a case.

The difficulty with balance-of-reason accounts, however, is that we want to know more than they offer to tell us. Suppose your reason for benefiting someone else is stronger than your reason for benefiting yourself, because the benefit to the other person would be larger and your helping her would not push you below some threshold of sufficient well-being. What more we want to know is whether you are morally *required* to sacrifice your own good for the benefit to the other person. If you are so required, then making the sacrifice is not supererogatory. But if there is what Hurka and Schubert (2012, p. 1) call an 'agent-favouring permission'—that is a permission for the agent to give his or her own good somewhat more weight than the same size good to someone else—then choosing not to sacrifice your own lesser good for the sake of someone else's somewhat greater good is morally permissible. And if choosing not to sacrifice your own lesser good for the sake of someone else's somewhat greater good is morally permissible, then choosing *to sacrifice* your own lesser good for the sake of someone else's somewhat greater good would be supererogatory.

Now suppose your reason for benefiting someone else is *not* stronger than your reason for benefiting yourself. This might be because the benefit to the other person would be *smaller* than the benefit to you that would come from not helping her. Or it might be, as Crisp envisions, because your helping her would push you below some threshold level of well-being. In such cases, the balance of reasons does not favour your helping her. That may be, but there is the further question whether you are morally *permitted* to help her.

The common-sense view is that you are morally permitted to pass up some benefit for yourself in order to benefit someone else even if the benefit to the other person would be *a bit* less than the benefit to you (Hurka and Schubert 2012, p. 2; Hooker 2013, p. 719; but compare Slote 1985, ch. 1; Crisp 2006, pp. 133–34). Admittedly, if the size of the gap between what was at stake for you and what was at stake for the other person is too large—e.g., you pass up giving yourself an extra decade of life for the sake of giving someone else an extra hour of life—you seem to have done something incompatible with self-respect. But if the size of the gap isn't too large, your giving someone else a benefit instead of giving yourself a bigger benefit is morally permissible, not wrong. Hurka and Schubert call this the 'agent-sacrificing permission'. What you've done in this example is downright admirable, because it's so unselfish. Here again we have a case of supererogation.

Let me address two possible objections here. One is the objection that a balance-of-reasons account *can* make sense of supererogation. The other is the objection that, although a balance-of-reasons account *cannot* make sense of supererogation, I am begging the question against Crisp by contending that a plausible ethical theory must make sense of supererogation.

Let me address the second objection first. I acknowledge that supererogation is a moral concept. It is the concept of action that goes above and beyond what moral duty requires. The concept of supererogation is parasitic on the concept of moral requirement. Crisp's ethical theory explicitly eschews moral concepts, and so *of course* this theory will reject the concept of supererogation, at least as it is normally understood.

If Crisp's theory does not aspire to accommodate supererogation, why think his theory should accommodate it? I accept that his theory is not guilty of inconsistency here. My point is that Crisp's theory gives up not only on the concepts of moral requirement and moral permission, but also on the concept of supererogation. Abandoning the concept of supererogation seems to me a huge drawback, given the importance of this category of moral acts. I would go so far as to say that, if a balance-of-reasons account is unable to allow for supererogation, this seems to me a fatal objection to such an account.

But the first objection I mentioned to my earlier conclusion was that a balance-of-reasons account *can* make sense of supererogation. In order to make sense of supererogation, a balance-of-reasons account might distinguish between *requiring* reasons and *good-making* reasons.[4] If one state-of-affairs contains more overall benefit than another state-of-affairs, this fact gives the first state-of-affairs more goodness of one kind than the other. Now take a case in which you could either benefit yourself a certain amount or benefit someone else a bit more. Here, the good-making reasons come down on the side of benefiting the other person, but the fact that the other person would benefit a bit more does not impose on you a requiring reason to choose what benefits the other person rather than what benefits you. Likewise, the idea might be that, when you could either benefit yourself a certain amount or benefit someone else a bit less, the good-making reasons come down on the side of benefiting yourself, but the fact that you would benefit a bit more does not impose on you a requiring reason not to benefit the other person instead.

The problem with this line of thought, in the present dialectical context, is that the line of thought reintroduces into a balance-of-reasons account a distinction between requiring reasons and good-making reasons. The concept of requiring reasons draws on the concept of requirements. Thus, a balance-of-reasons account that is trying to paint a picture of ethics without employing moral concepts will not have available on its palette a distinction between requiring reasons and good-making reasons.

In this section, I have contended that at least sometimes we want to know more than what the balance of de-moralized reasons favours. Sometimes, we want to know whether declining to do what the balance of such reasons favours would be wrong. In cases in which declining to do what the balance of such reasons favours would *not* be wrong, space for supererogation opens up. I've suggested that some attributions of moral permissibility and some attributions of supererogation cannot be reduced to judgements about balances of de-moralized reasons.

[4] Joshua Gert distinguished between requiring reasons and justificatory reasons (2004, especially ch. 5). What I am calling good-making reasons are merely one kind of justificatory reasons.

Crisp might be taken to have anticipated this line of objection: 'A second objection … to my proposal is that it makes it impossible in certain cases for us to say what we want to say. All I could say about Hitler, for example, might be that he had a reason (albeit a very strong reason) not to do many of the things he did.' (p. 27) Crisp accepts that, at the level of philosophical reflection on ethics, all one needs to say can indeed be expressed in terms of de-moralised reasons of greater or lesser strength. But he stresses that his contention does not deny moral concepts to everyday ethical language and thought.

Undeniably, everyday ethical thought often zeros in on questions about what is wrong, about what is permissible instead of wrong, or about what is supererogatory instead of morally required. But Crisp's de-moralizing campaign targets philosophical reflection on ethics, ethical theorizing, rather than everyday moral thought. I accept that, for the sake of explaining more on the basis of fewer postulates, ethical theorizing can try to start out with far fewer concepts than are used in everyday moral thought. Nevertheless, because of the arguments above, I do not think wrongness, permissibility, and supererogation can be adequately explained in terms of the balance of de-moralized reasons. Moreover, I think ethical theorizing is continuous with everyday moral thought in taking questions about what is wrong, permissible, or supererogatory to be immensely important.

## 4 Crisp's Argument from an Idealized Community of Rational Beings

Crisp's argument from an idealized community of rational beings is as follows:

> Imagine a community of rational beings who are benevolent to the point where there is no need for the sanctions of law and morality to guide their behaviour, though they may perhaps have certain rules for the purposes of coordination and cooperation. If they know nothing of law and morality, it does not seem that they are missing something—in particular, the idea that certain actions are wrong, and that this wrongness provides reasons for action…. (p. 17)

Crisp presumably means 'there is no need for the sanctions of law and morality to guide their behaviour' because these rational beings are equipped with enough benevolence plus rules for the purposes of coordination and cooperation to ensure that *these beings always behave in morally correct ways, even though they had no ideas about wrongness*. Is it plausible that rational beings with enough benevolence plus rules for the purposes of coordination and cooperation would always behave in morally correct ways and always react to others' behaviour in morally correct ways?

Before I start addressing Crisp's argument from an idealized community of rational beings, I note that this argument sits uncomfortably alongside the distinction between de-moralizing as the programme of expunging moral concepts from philosophical reflection on ethics and de-moralizing as a proposal for everyday moral thought. On the one hand, Crisp's argument from an idealized community of rational beings concerns the everyday moral practice of these rational beings. On the other hand, I identified his programme as de-moralizing philosophical reflection on ethics, not de-moralizing everyday moral practice. I guess his argument from an idealized community of rational beings has as a premise the conditional that, if nothing would be lost from de-moralizing the everyday moral decision

making of an idealized community of rational beings, then we have nothing to lose from de-moralizing philosophical reflection on ethics. I am not sure whether to accept this conditional. But rather than assess it, I will argue that its antecedent is not true.

The next question to pose is how much benevolence these rational beings are supposed to have. If the answer is impartial benevolence, then I cannot see how these beings will recognise the agent-favouring moral permissibility of declining to pass up a benefit for oneself for the sake of a slightly larger benefit's going to someone else. Nor will these beings recognise the agent-sacrificing moral permissibility, indeed admirability, of passing up a greater benefit for oneself for the sake of giving a somewhat smaller benefit to someone else. Having already used these examples above, I will not say more about them here.

The passage I quoted implies that all the sensitivities a moral agent needs can be reduced to benevolence and commitment to rules for the purposes of coordination and cooperation. Among the moral concepts that Crisp's de-moralizing project explicitly eschews are 'fairness' and 'justice' (p. 25). But given that concerns for fairness, justice, rights, and desert can conflict with benevolence, can a well-functioning moral agent do without these concerns?

Fairness is a multi-dimensional concept. One dimension consists in unbiased application of criteria and rules. A second dimension concerns the content of rules. Rules are fair only if they make all and only the appropriate distinctions. Which distinctions rules should make will depend on which social practice is structured by the rules. Stepping back from individual social practices, we can ask whether a society's whole ensemble of practices is fair.

One view is that a society is fair if and only if and because its ensemble of practices maximizes expected aggregate utility, impartially calculated. Perhaps, if this is the correct view about whether a society is fair, then agents who are impartially benevolent and have rules for the purposes of coordination and cooperation might not need an additional concern for fairness, since fairness would lead them to the same conclusions about what practices are justified as impartial benevolence would.

Suppose that the ensemble of practices that maximizes expected aggregate utility impartially calculated would leave some people very much better off than others. If there is an alternative ensemble of practices that would produce almost as much expected aggregate utility as the first one would but would leave the worst-off less badly off, then many people would think fairness would favour this second ensemble of practices. If impartial benevolence focuses exclusively on expected aggregate utility while fairness takes into account how badly off the worst-off are left, then agents who are impartially benevolent and have rules for the purposes of coordination and cooperation will *not* favour the same ensemble of practices as will agents who not only are benevolent and have rules for the purposes of coordination and cooperation but also care about fairness.

Crisp might understand impartial benevolence such that it would give extra weight to the well-being of those below some threshold, which is different from a concern for the worst-off since the worst-off might be above this threshold (see his ch. 6). I admit that construing impartial benevolence as including extra concern for those below some threshold partly undercuts an argument that juxtaposes impartial benevolence with a concern for fairness.

Turn now to arguments about justice, rights, and desert. However, one leading theory of justice is Rawls's 'justice as fairness' (1971). Distinguishing between fairness and justice is thus controversial. Furthermore, another leading theory of justice is that justice consists in according to people what they have rights to and what they deserve. But the relations between fairness, justice, rights and desert are too big a topic to address here.

Luckily, we do not need a convincing account of the relations between fairness, justice, rights, and desert in order to make points relevant to Crisp's argument. Even if the *general practice* of giving people what they have rights to and what they deserve maximizes expected aggregate well-being, *in many individual cases* giving people what they deserve or what they have rights to will not maximize expected aggregate well-being, and thus is not what someone motivated purely by impartial benevolence would do. That is true whether we understand impartial benevolence as a utilitarian would or understand impartial benevolence as including extra concern for the well-being of those below some threshold.

Would commitment to rules for the purposes of cooperation and coordination ensure that agents always choose in accordance with others' rights? Rules for coordination and cooperation underpin many rights. But it is hard to believe that *all* of what people have rights to is determined by applying rules which exist to solve coordination and cooperation problems.

As an example, consider the rights that your family and friends have to your loyalty or partiality when you are deciding how to allocate your own resources (cf. Crisp, pp. 142–44). Suppose you could spend your Sunday afternoon helping either your friend or someone who is not especially connected to you. Suppose also that the benefit to your friend would be a bit less than to the other person, and that both your friend and the other person are above a threshold of sufficient well-being. Thus, *impartial* benevolence, whether construed in a utilitarian way or in Crisp's way, would lead you to help the other person rather than your friend. But most people have the intuition that you owe your friend at least some degree of priority over someone who isn't you friend when you are deciding how to allocate your own time and energy. Corresponding to your duty of loyalty to your friend is a right your friend has to at least a degree of priority over strangers when you are deciding what to do with your own resources. But impartial benevolence in itself wouldn't recognise this duty and the corresponding right.

Crisp seems to think that rational beings with the right amount of benevolence and an unswerving disposition to conform with rules for cooperation and coordination could behave correctly even if they had no thoughts about morality. I have contended that rational beings with the right amount of benevolence and an unswerving disposition to conform with rules for cooperation and coordination would also need to be sensitive to considerations of fairness and rights (such as the right to loyalty). I turn now to desert.

Suppose that Jack and Jill are engaged in the same enterprise, e.g., writing novels, painting portraits, or designing apps. And suppose that what Jill produces is much better than what Jack produces, and yet Jack receives more praise. If Jack had improved more quickly than Jill or had more obstacles to overcome than Jill, then his greater improvement or greater perseverance could intelligibly be praised more than Jill's improvement or perseverance. Such praise for greater improvement or greater perseverance is not for having produced something better than she produced. But let us assume the praise is indeed focused on the things produced, not on something about the process of producing them. If so, then it is unjust for Jill's product to receive less praise than Jack's. Her product is better, and so, considered by itself, deserves higher praise.

One sensitivity that agents will need in order never to act wrongly is a sensitivity to desert. This sensitivity cannot be the same as benevolence, since they often pull in opposite directions. Nor can a sensitivity to desert be entirely reduced to commitment to rules for coordination and cooperation, since desert sometimes comes into play outside those contexts. That seems to me to be true of the example about praise for Jack and Jill.

The very idea that people deserve that a certain reactive attitude be targeted on them appeals to concepts that de-moralizers propose we jettison. For example, you deserve praise for working yourself to exhaustion helping others and doing far more to help than others have done. The extent to which you made sacrifices to help others was supererogatory, and supererogatory acts are praiseworthy. Another example is that the indignation I deserve for stealing the credit that Samantha should have gotten depends on the fact that I was morally wrong to steal her credit.

Remember that we are addressing Crisp's suggestion that rational beings with the right amount of benevolence and an unswerving disposition to conform with rules for cooperation and coordination might behave correctly even if they had no thoughts about morality. I have suggested that such beings might sometimes behave wrongly or have inappropriate attitudes, because they lack concepts of and sensitivities to fairness, rights, and desert. If some of these rational beings would sometimes behave wrongly, they would sometimes deserve indignation. And they might sometimes deserve punishment.

There is a long-standing argument that punishment is *by definition* harsh treatment not only inflicted on people for having done acts that are thought to be wrong and but also expressing condemnation of those acts as wrong. If the rational beings that Crisp describes have no concept of wrongness, can they have a decent conception of justified punishment (Hooker 2006; McElwee 2010)?

Some act-consequentialists have argued that justified punishment is not tethered to wrongness. They have pointed to examples where inflicting punishment on the innocent somehow brings about very good consequences (typically by appeasing angry mobs or deterring others from crime). These examples typically stipulate that the punished person's innocence is kept secret, as is the principle acted upon, i.e., the principle that *the innocent should be punished when doing so will produce the best consequences*.

The idea that punishment might sometimes be *appropriately* inflicted on the innocent is highly revisionary. So is the idea that principles might appropriately be kept secret. These ideas discard central tenants of our conception of how society should be run—such as the tenant that punishment should be inflicted 'only on individuals who have had a fair opportunity to avoid being subject to it' (Scanlon 2013, pp. 101–16; 2018, ch. 8). Perhaps, Crisp's rational beings with the right amount of benevolence and rules for cooperation and coordination would formulate some other replacement for our concept of justified punishment. Whatever the replacement is, however, if this replacement has to work without reference to the concept of wrongness, it would be very different from our concept.

## 5 Conclusions

The ethical ideas in which Crisp has greatest confidence are the idea that there are non-derivative reasons to promote one's own well-being and the idea that there are non-derivative reasons to promote impartial well-being. He thus thinks ethical theorizing should take these dualistic reasons to promote well-being as the normative foundation on which to build the rest of ethics. For him, that there are these non-derivative reasons to promote well-being is also epistemically foundational, the foundation from which we work out what else we know in ethics. He has too little confidence in the concepts and rules of common-sense

morality to think that whether an ethical theory is acceptable or not can turn on whether its implications accord with common-sense morality.

In contrast, I have greater confidence in the rules of common-sense morality and ideas about supererogation and justified punishment than I do in any proposed single unifying principle or in Crisp's dualism of egoism and altruism. Thus, whether a proposed ethical theory seems to me acceptable turns on, among other things, whether its implications fit with common-sense ideas about what is morally required, what is morally permissible, and what is supererogatory. Thus, Crisp and I disagree about what is epistemically primary in ethics.

Our disagreement in the epistemology of ethics does not preclude substantial agreement in normative theory. We agree that the deepest explanations in ethics will not rest content with the rules of common-sense morality, but instead push beyond these to more basic normative foundations. And we both think that possible moral systems (possible positive moralities) should be assessed exclusively in terms of consequences for well-being plus, in his view, sufficientarian qualifications (pp. 158–62) and, in my view, prioritarian weighting (Hooker 2000, pp. 55–64).

**Acknowledgements** This paper was presented at a 2019 workshop organised by Tyler Paytas and Richard Rowland and held at the Rome campus of the Australasian Catholic University. I am grateful for David Killoren, Roger Crisp, Alastair Norcross, Thomas Schmidt, David Plunket, Richard Rowland, Marcia Baron, Julia Driver, Elinor Mason, Neil Shinhabatu, Stephen Finlay, Piers Rawling, Terry Irwin, and Tyler Paytas for helpful comments on my paper at the workshop. Later, the paper's arguments were the focus of a discussion group at the University of Reading, and there I received insightful questions and advice from David Oderberg, Charlotte Newey, Emma Borg, Nat Hanson, Max de Gaynesford, Luke Elson, and Philip Stratton-Lake. Finally, comments from two reviewers for this journal led to further improvements to the paper.

## Declarations

The author has no relevant financial or non-financial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

## References

Crisp R (2006) *Reasons and Goodness*. Clarendon Press, Oxford

Gert J (2004) *Brute Rationality*. Cambridge University Press, Cambridge

Hooker B (2000) *Ideal Code, Real World: A Rule-consequentialist Theory of Morality*. Clarendon Press, Oxford

Hooker B (2006) 'Right, Wrong, and Rule-consequentialism', in West H (ed.), *Blackwell Guide to Mill's Utilitarianism*. Blackwell Publishers, Malden, MA, pp. 233–48

Hooker B (2009) 'The Demandingness Objection', in Chappell T (ed.), *The Problem of Moral Demandingness*. Palgrave Publishers, pp. 148–62

Hooker B (2013) 'Egoism, Partiality, and Impartiality', in Crisp R (ed.), *Oxford Handbook on the History of Ethics*. Oxford University Press, Oxford, pp. 710–28

Hooker B (2018) 'Moral Theory and Its Role in Everyday Moral Thought and Action', in Zimmerman A, Jones K, Timmons M (eds.), *The Routledge Handbook on Moral Epistemology*. Routledge, New York, pp. 387–400
Hurka T (2014) *British Ethical Theorists from Sidgwick to Ewing*. Oxford University Press, Oxford
Hurka T, Schubert E (2012) 'Permissions to Do Less than the Best: A Moving Band'. *Oxf Stud Normative Ethics* 2:1–27
McElwee B (2010) 'Should We De-Moralize Ethical Theory?'. *Ratio* 23:308–320
Parfit D (2017) *On What Matters*, vol 3. Oxford University Press, Oxford
Rawls J (1971) *A Theory of Justice*. Harvard University Press, Cambridge MA
Scheffler S (1982) *The Rejection of Consequentialism*. Clarendon Press, Oxford
Scanlon T (2013) 'Giving Desert its Due'. *Philosophical Explorations* 16:101–116
Scanlon T (2018) *Why Does Inequality Matter?* Oxford: Oxford University Press
Slote M (1985) *Common-Sense Morality and Consequentialism*. Routledge and Kegan Paul, London
Stratton-Lake P (2016) 'Intuition, Self-Evidence, and Understanding'. *Oxf Stud Metaethics* 11:28–44

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.