ARE THERE REASONS TO ACT MORALLY?[*]

Michael Smith

1. The diversity of normative standpoints

Let's say that we appraise people's conduct from a *normative* standpoint whenever we compare what they do with the actions that they are obliged to do, permitted to do, and prohibited from doing, or more generally with the actions they have reason to do, no reason not to do, and reason not to do. Such appraisals are ubiquitous. Indeed, it is hard to imagine any sort of social life without them. But that is not to say that such appraisals are well understood.

The first and most obvious thing to say about normative appraisal is that there are many different normative standpoints that issue in such appraisals. Some are local and idiosyncratic, as when you are a member of a club whose rules tell you what you are obliged to do, permitted to do, and prohibited from doing, as a club member. These rules are local and idiosyncratic because club rules vary from club to club depending on the different interests of those who join. Other normative standpoints are more global, but still localized to certain geographical regions, as when you live in a state and the laws of that state tell you what you are obliged, permitted, and prohibited from doing as someone who is within the state's borders. The laws of states are also idiosyncratic because the laws within a state depend on the interests of some group within each state—in a democracy these are the interests of everyone eligible to vote, in a plutocracy they are the interests of the wealthy, in a dictatorship they are interests of the dictator, and so on—and these interests will differ from state to state.

There are also normative standpoints from which we can evaluate what one human being can do to another without regard to which state they live in. Such a standpoint was expressed in the Universal Declaration of Human Rights, adopted by the United Nations in 1948. However, as the name makes clear, the appraisals it permits are still idiosyncratic, as the content of the declaration reflects the interests of the voting members of the UN in 1948, and they are also highly restricted, as only the actions of human beings that affect other human beings fall within its scope. The Declaration has no application at all to the actions of human beings that affect non-human animals, not even those non-human animals with remarkably sophisticated psychologies like higher apes and dolphins, and nor does it apply to the actions of extra-terrestrials, if there are any, not even those extraterrestrials, again if there are any, who are rational agents in exactly the same sense in which humans are rational agents. (I note in passing that there is movement among animal rights activists to have the UN adopt a universal declaration concerning the rights of Great Apes.)

Observations about how many different normative standpoints there are and the differences between them prompt an obvious question. Are there any normative standpoints that are thoroughly global and non-idiosyncratic? The short answer is that there are two such standpoints.

---

[*] This is one of a series of lectures given at Nanjing Normal University and Shandong University in November 2018. The lectures are best read in the following order: "From Knowledge of Our Existence to Knowledge of Normative Reasons", "Are There Reasons to Act Morally?", and "Twenty-First Century Non-Naturalism". I am extremely grateful to Professor Zhen CHEN and Professor Tongli WU for making my visit to China both possible and so enjoyable, and to all those who asked questions at the lectures for giving me so much to think about.

The longer answer concerns the relationship between these two standpoints. The aim in what follows is to give that longer answer.

2. The standpoint of reason

Normative standpoints that are global and non-idiosyncratic would have to be normative standpoints from which we appraise the actions of every possible rational agent—humans and extraterrestrials with the same capacities for rational agency as humans, if there are any—simply in virtue of the fact that they are rational agents. There is obviously at least one such standpoint because we appraise the actions of every possible rational agent, simply in virtue of being a rational agent, from the standpoint of *reason* itself. Let me therefore begin by saying a little about the standpoint of reason.

What it is to be a rational agent is to be someone with the capacity to form beliefs about how the world is in the light of the available evidence, and, if suitably situated, to gain knowledge of the world, and it is also be someone with the capacity to realize their desires about how the world is to be. This means that we can appraise the actions of rational agents by the standards internal to knowledge-acquisition and desire-realization, that is, by the standards of theoretical reason and practical reason. These two together constitute the standards by which appraise agents from the standpoint of reason. Though talk of appraising agents by the standards theoretical reason and practical reason might suggest that we can only criticize those who are capable of high-level thinking and complex planning, it is important to realize that these standards allow us to appraise even the actions of quite unsophisticated agents.

Imagine a very simple agent, a fish of some sort, in an environment that has been rigged up to make it seem to the fish that there is food in a certain place where there is none. The fish goes to that place and engages in food-consuming behavior, but to no avail. The fish's actions, being based on a desire for food and a misrepresentation of the location of food in its environment, thus fail in their own terms. The fish doesn't get any food. Though the fish isn't capable of high-level thinking or complex planning, its representation of the world is still defective from the standpoint of reason because the actions its representation leads it to perform fail reliably to satisfy the desires that brought them about. Reason itself thus tells us what the representations that play a role in the production of action have to be like in order to play the role that they play. Such representations have to be reliably apt for guiding us so as to realize our desires, and this in turns requires that they be accurate or true.

When we switch to examples of agents who are capable of high-level thinking and complex planning, we can imagine more complex failures, failures that go beyond misrepresentation. My beliefs, true or false, may be formed not just in light of the available evidence, but also in light of the way I respond to that evidence. Reason itself tells us that there are failures of belief in virtue of the way I respond. My beliefs can be *theoretically irrational* because I fail to attend to all of the available evidence, or attend to it, but in the wrong way. Whether or not such beliefs misrepresent how things are, when they combine with my desires to produce an action, the actions I perform on the basis of such beliefs are irrational in a corresponding way. Moreover this is so even if by some fluke they happen to be true and so lead me to act in a way that realizes my desires. They are irrational because they are still not such as to *reliably* combine with desires so as to produce actions that satisfy those desires.

There are also ways in which such agents can go wrong in their desire-formation. Imagine that I have two desires, a weaker desire to have pleasant experiences and a stronger desire to be

healthy, and that I am confronted by a delicious looking and smelling dish of food that I know to be very unhealthy. My options are to eat or not to eat. In such circumstances it is easy to imagine that, even though I don't misrepresent the effects of my options, and even though what I believe about my options is supported by the available evidence, I get so carried away by the smell and look of the delicious looking food that I eat it anyway, notwithstanding the fact that I know I shouldn't because of its bad effects on my health. Here we have a normative assessment of my action—I eat the food even though I shouldn't—that isn't explained by a misrepresentation, or by the theoretical irrationality of some belief I have. What explains this assessment?

Norms of *instrumental rationality*—that is, the norms that govern what we do in the light of the fact that we desire various things more or less strongly and have many different beliefs about how these different desires are to be realized, beliefs that we hold with different levels of confidence—are themselves norms to which we may conform either locally or globally. Whenever we act, we must conform to such norms locally, because the fact that we act shows that we have managed to put some desire together with some belief we have about how that desire is to be realized. In the example just given, I pick up my knife and fork and transfer the delicious looking food from the plate to my mouth, I chew and swallow, all with the aim of getting the desired pleasure. But my being locally instrumentally rational in this way is consistent with my being more globally instrumentally irrational.

In order to see this, imagine that in the case just described my weaker desire for pleasure hooks up with my beliefs about how to get pleasure in such a way as to act accordingly, but that, because of my preoccupation with the look and the smell of the delicious food, my stronger desire to be healthy does not hook up with my beliefs about what I need to do in order to be healthy. This is also a case in which I act in a way that I shouldn't. I act in a way that I shouldn't because, even though I am locally instrumentally rational, I am more globally instrumentally irrational. Agents who are locally instrumentally rational may therefore succeed in satisfying some desire that they have, but if they are globally instrumentally irrational, they will fail to satisfy their desires *overall*: that is, they fail to satisfy their different desires in proportion to their strength.

I said earlier that the standpoint of reason is global and non-idiosyncratic. It is global because we can appraise the conduct of every possible rational agent from that standpoint. Humans and extraterrestrials who are rational agents in exactly the same sense in which humans are rational agents are thus all subject to appraisal from the standpoint of reason. It is also non-idiosyncratic because there is just one standpoint of reason, not multiple such standpoints. This one standpoint tells us how the desires and beliefs that produce actions have to be if they are to meet the standards internal to action itself. The standards internal to action itself are, as we have seen, the standards of theoretical and practical reason. When we criticize agents' actions from the standpoint of reason, we do so because they act in the way they do as a result of misrepresenting the world (this is a defect by the standards of theoretical reason), or as a result of having formed their beliefs about the world improperly in response to the available evidence (this is also a defect by the standards of theoretical reason), or as a result of their global instrumental irrationality (this is a defect by the standards of practical reason).

Moreover, a further kind of criticism from the standpoint of reason may also be available, given that these are available. Many agents who act in a way that is criticizable from the standpoint of reason in one of the ways just mentioned are in a position to know that they're acting in that way when they do, and, even before they act in that way, many can anticipate that they will act in that

way if they don't exercise self-control. An exercise of self-control can thus prevent such agents from being criticizable. Agents who have the capacity for self-control but don't exercise it can therefore be criticized for their failure to do so (this is a further defect by the standards of practical reason).

It is worth pausing to ask what exactly it might mean for agents to have and exercise the capacity for self-control. There are many ways in which agents in fact exercise self-control. One familiar way in which humans exercise self-control is by imagining things that cause their motivational profile to change from one that is criticizable from the standpoint of reason to one that isn't. Think again about a situation in which the smell and look of delicious food causes me to be instrumentally irrational so that I desire most to pursue the means to satisfying my weaker desire for pleasure rather than my stronger desire to be healthy. This is a paradigmatic situation in which the exercise of self-control is called for. Suppose that I have promised my mother not to give in to the temptation of unhealthy food, and that if I were to imagine the disappointment on her face if she knew that I was going to eat unhealthily, my motivations would transform so that I choose the healthy option. In this situation, my capacity to exercise self-control amounts to my capacity to imagine the disappointment on my mother's face at the crucial moment.

Importantly, what it is to exercise self-control therefore turns out to be a largely empirical matter. It will depend on which psychological triggers agents can pull to transform their motivational profile from one that is criticizable from the standpoint of reason to one that isn't, and these psychological triggers will differ from agent to agent. In typical humans, imagining the look of disappointment on one's mother's face is such a psychological trigger. But not all humans are typical, and such an imaginative feat might not even be possible for extraterrestrial agents who come into existence and develop without the involvement of a mother with whom they develop an intimate bond. The upshot is that self-control for extraterrestrials might well take a form that we humans find difficult even to understand.

Whatever it is for a rational agent to possess the capacity for self-control, possession of that capacity will assume an especially important role in the lives of those who have it, humans and extra-terrestrials alike. This is because those who possess the capacity for self-control can elude criticism from the standpoint of reason in a more modally robust way than those who lack it. All is not lost if such agents find themselves vulnerable to acting in a way that is criticizable from the standpoint of reason, as they can gain knowledge of their vulnerability and then take steps to ensure that they don't act in that way. This in turn means that those with self-control will be criticizable from the standpoint of reason if they lack knowledge of more than just means-to-ends. *Self*-knowledge—a rational agent's knowledge of what they would have to do in order to elude criticism from the standpoint of reason, knowledge that they are vulnerable to acting in a way that is so criticizable, and knowledge of the psychological triggers that they can pull by way of remedy—will also be of crucial significance. We will return to this point presently.

Let's sum up the story so far. We have seen that agents' actions are criticizable from the standpoint of reason when they act as a result of misrepresenting the world, or having formed their beliefs about the world—including themselves—improperly in response to the available evidence (these are defects by the standards of theoretical reason), or their global instrumental irrationality, or their failure to exercise self-control when they had the capacity to exercise it (these are defects by the standards of practical reason). But is this list exhaustive? In particular, does the standpoint of reason have anything to say either for or against agents having the intrinsic desires that lie at the very source of their actions, or is it entirely permissive? In

answering this question, we have to keep in mind some obvious empirical facts and one uncontroversial normative truth.

The obvious empirical facts are that different people have different intrinsic desires; that the same person can have different intrinsic desires at different times in their life; that when two people have the same intrinsic desires, they can have those desires with very different strengths; and that when someone has the same intrinsic desires over the course of their lifetime, those desires can vary in strength from time to time. The uncontroversial normative truth is that such variations in intrinsic desire are often permissible from the standpoint of reason. To give just one example, consider the intrinsic desire to eat a certain flavor of ice-cream. You can have this intrinsic desire or lack it; if you have it at one time you can give it up at another; when you have the intrinsic desire, it can be strong or weak; and, most importantly, all of this is rationally permissible. The standpoint of reason is thus entirely permissive when it comes to intrinsic desires like the desire to eat a certain flavor of ice-cream.

These empirical and normative facts force us to clarify the question. Are there some things that all rational agents should intrinsically desire to do, or some things that they should not intrinsically desire to do, simply in virtue of being rational agents? Note that we can ask the same question in terms of agents' reasons for action, given the platitudinous connection between agents acting in ways that elude criticism from the standpoint of reason and their doing what they have reason to do. Do agents have reasons to satisfy the intrinsic desires that they happen to have, in proportion to their strength, no matter what their content, or are there some things that they have reasons to do quite independently of the intrinsic desires that they happen to have?

Suppose that there are such intrinsic desires, and hence that there are such reasons for action. In that case there must be some further dimension of assessment of an agent’s beliefs and desires from the standpoint of reason beyond those already mentioned: that is, beyond misrepresentation, beyond a failure to believe what's supported by the available evidence, beyond global instrumental irrationality, and beyond the failure to exercise self-control. There must be some such further dimension because these suggest that no intrinsic desires are either required or forbidden from the standpoint of reason.

3. The standpoint of morality

Before trying to figure out what this further dimension of assessment might be, we need to consider another global and non-idiosyncratic normative standpoint, one which we haven't talked about so far. This is because we must take care not to slide from appraisals from the standpoint about which we are asking these questions, which is the standpoint of *reason*, to appraisals from this other standpoint, which is the normative standpoint of *morality*.

Consider an action like grievously harming someone. Imagine that action being performed by a fish: a shark attacks a swimmer. Though what the shark did was awful, we don't appraise the shark's action in moral terms. Sharks aren't subject to the moral prohibition on harming people, and the reason why would seem to be clear. Agents can only be subject to a moral prohibition on harming people if they can understand that what they are doing when they harm someone is harming them. Sharks can't understand that their actions have these effects, so they are not subject to the moral prohibition.

Now imagine the action of grievously harming someone being performed by a young child or someone with impulse control problems. Though they may be able to understand that what they

are doing when they harm someone is harming them—and hence though they may be subject to the moral prohibition, just like normal human adults—they might not have the capacity to understand that it is morally impermissible to harm people (the child), or the capacity for self-control required to get themselves to refrain from doing what they're in a position to know it would be morally impermissible to do when they're not antecedently inclined to refrain from doing so (the person with impulse control problems). As such, even though may both be subject to the moral prohibition, young children and those with impulse control problems might be excused for violating it. It might not be legitimate to hold them responsible.

Now imagine the action of grievously harming someone being performed by normal human adults who understand that what they're doing when they harm someone is harming them, who are in a position to know that such behavior is morally prohibited, and who have the requisite capacity for self-control that would get them to abide by the prohibition if they are not antecedently inclined to abide by it. They are subject to the moral prohibition and they have no excuse if they knowingly or negligently violate it. We can therefore legitimately hold them responsible. In other words, it may be morally permissible, or even morally required, for us to blame them for their violation, where we can think of 'blame' as the name for whatever the appropriate response is, from the moral standpoint, to a violation of the moral prohibition on harming people in the circumstances in question, something that could vary depending on the circumstances. It might amount to reprimanding the person who caused the harm, or requiring that person to compensate those he harmed, or cutting them off from certain sorts of social relations, or even just to keeping track of their violation with a view to taking one of the measures already mentioned if a pattern of similar violations continues.

Here, then, we have another standpoint, the standpoint of morality, that is also non-idiosyncratic and almost thoroughly global. I say that it is almost thoroughly global because it is a standpoint from which we can appraise the conduct of all possible rational agents *who can understand the effects of their conduct on other rational agents*. The actions of humans who meet this condition are thus subject to moral appraisal, but so too are the actions of extra-terrestrials, if there are any who meet this condition, and so too are the actions of merely possible agents, characters in fictions, mythological beings, and so on. The moral standpoint is also non-idiosyncratic because, unlike the rules of a club or the laws of a state or the Universal Declaration of Human Rights, what's morally obligatory, permitted, and prohibited isn't fixed by the interests of some group. We will have more to say about what fixes the facts about what's morally obligatory, permitted, and prohibited below.

Of course, to say that the moral standpoint is global and non-idiosyncratic is not to deny that people can radically disagree with each other about what is obligatory, permitted, and prohibited from the moral standpoint. They can and do disagree with each other. Moreover, to say that these facts are global and non-idiosyncratic is consistent with there being great circumstantial variation in what is obligatory, permitted, and prohibited from the moral standpoint. To give just one example, what counts as harming a rational agent will depend on the nature of that agent's body. The moral prohibition on harming thus entails a prohibition on punching and slashing human beings, as these are ways in which the bodies of human beings can be damaged, and one kind of harm is linked to bodily damage. But there may be no corresponding prohibitions on punching or slashing extraterrestrials or imaginary agents whose bodies can't be damaged by fists or knives.

We noted a short while ago that when we criticize the actions of a rational agent from the standpoint of reason, we do so because they act in the way they do as a result of misrepresenting

the world (including themselves), or having formed their beliefs about the world (including themselves) improperly in response to the available evidence, or their global instrumental irrationality, or their possessing but failing to exercise the capacity for self-control. The question we asked is whether this list is exhaustive. In particular, we asked whether we can criticize the intrinsic desires that lie at the source their actions from the standpoint of reason. Are there certain things that we should intrinsically desire to do, or should not intrinsically desire to do, or is it permissible to intrinsically desire what we like? Are all intrinsic desires like the desire for ice-cream?

We can now see that in answering this question we must take special care not to slide from appraisals from the standpoint of reason to appraisals from the standpoint of morality. Since it is morally impermissible not just to harm people, but also to have intrinsic desires that would lead one to harm people, it follows that there are certain intrinsic desires that one shouldn't have from the standpoint of *morality*. But that isn't the question we are asking. We are asking whether there are certain intrinsic desires that one shouldn't have from the standpoint of *reason*. We can also clarify the question in terms of reasons for action. The question isn't whether harming people is morally impermissible, but whether there is a reason not to harm people. It is to that question that we turn next.

4. Kant vs Hume on the relationship between the standpoints of morality and reason

Many take this question to have an obvious answer, as they think that facts about what's obligatory, permissible, and prohibited from the standpoint of morality are themselves fixed by facts about what's obligatory, permissible, and prohibited from the standpoint of reason. In Western philosophy, disagreement on this issue plays out in the work of Hume (*Treatise* 1740) and Kant (*Groundwork* 1786).

While Hume had quite conventional views about what is morally required, permitted, and prohibited, he thought that reason puts no constraints at all on the ends that we can pursue. As he infamously puts it:

> Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an Indian or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than for the latter. (Hume *Treatise* 2.3.3.6)

Hume comes to this conclusion because he thinks that to criticize agents' actions is to criticize them for failing to pursue the ends that will *in fact* satisfy their intrinsic desires. In his view, agents have reasons to satisfy their intrinsic desires no matter what their content.

Kant's view couldn't be more different. Many actions that in fact satisfy the intrinsic desires that agents happen to have are actions that they have reasons not to perform. This is because, as he sees things, moral requirements are themselves "categorical" requirements of reason, which is to say that they are reasons to act that agents have simply in virtue of their being rational and so quite independently of their antecedent desires and interests. Kant calls the motive to perform such actions "respect for the moral law", but in our terms this motive is an intrinsic desire whose content picks out the actions that are morally required. In Kant's view this motive—that is, these intrinsic desires—are themselves required by reason.

How are we to adjudicate this disagreement? It is easy to see the attraction of Kant's view. Kant thinks that there is what we might call a *morality-reason-responsibility nexus*. As we have seen, if there is a moral prohibition on causing harm then this prohibition applies to all rational beings who can understand that their actions cause harm, humans and extraterrestrials alike, and those who violate this moral prohibition can legitimately be held responsible for having so acted if they have no excuse, where an excuse amounts to either an incapacity to understand that harming another is morally prohibited, or a failure to possess the powers of self-control required to get yourself to abide by the moral requirement if you aren't antecedently inclined to do so. But for the moral prohibition to have this scope, and for these to be the only excuses, it seems that it must be reasonable to expect those who meet these conditions not to cause harm, and this in turn suggests that the powers of reason must themselves be sufficient for agents to grasp the moral prohibition and to get themselves to act accordingly. Moral knowledge and self-control must be available to humans and extra-terrestrials alike simply in virtue of the fact that they are rational agents with the requisite capacities.

If Kant is right then what we understand, when we understand moral prohibitions, is that those rational agents who violate such prohibitions, absent an excuse, are criticizable from the standpoint of reason, and the powers of self-control must be powers to get ourselves to desire to do, and then to do, not just what it would globally instrumentally rational for us to do, but all those things that we would be criticizable from the standpoint of reason for failing to do, given that we possess such powers. In other words, if there is indeed a morality-reason-responsibility nexus then, contrary to Hume, the existence of a moral prohibition on causing harm itself entails that having an intrinsic desire to cause harm that leads all the way to action is a violation of the standards of practical reason. The existence of the moral prohibition entails a reason not to cause harm.

Having said this, it is also easy to see the attraction of Hume's view. As we have seen, when we criticize the actions of a rational agent from the standpoint of reason we do so either because their actions are based on a misrepresentation of the world (including themselves), or because they have formed the beliefs that led them to act improperly in the light of the available evidence, or because their actions are in part explained by their global instrumental irrationality, or because they possess but fail to exercise the capacity for self-control. Since none of these dimensions of criticism license us to criticize rational agents who have an intrinsic desire to cause harm that leads all the way to action, even if Kant is right that there is a morality-reason-responsibility nexus, that doesn't provide us with any insight at all into what the powers of reason are that suffice for knowledge of the moral prohibition on causing harm, or for getting ourselves to conform to such a prohibition.

Of course, we could give both Hume and Kant their due in this disagreement by agreeing with Kant that there is a morality-reason-responsibility nexus, but then insisting that this is merely a *conceptual* truth: If there is a moral prohibition on causing harm, then there is a reason to cause harm, from which it follows that having an intrinsic desire to cause harm that leads all the way to action is rationally impermissible. This would be to agree that the concept of a moral prohibition is, inter alia, the concept of an action that we would be criticizable from the standpoint of reason for performing independently of our antecedent desires and interests. But since this conceptual truth about moral prohibitions is consistent with our learning from Hume that having an intrinsic desire to cause harm that leads all the way to action is not rationally impermissible, the proper conclusion to draw might be that nothing falls under the concept of a moral prohibition. There

are no reasons to act as morality requires. In other words, the proper conclusion to draw might be an Error Theory about morality in the spirit of John Mackie (1977).

But though we could give both Hume and Kant their due in this way, it seems to me that it would be a mistake to do so. Kant isn't just right about the concept of a moral prohibition, he is also right that there are moral prohibitions: the concept is instantiated. In order to see that this is so, we need to think more carefully about the ways in which we can criticize those who have the capacity to have knowledge of the world in which they live and realize their desires in that world. *Ideal* agents—that is, agents who elude *all* criticism from the standpoint of reason—must robustly have and exercise these capacities to the greatest degree possible. As we have already seen, the fact that agents have and exercise the capacity for self-control contributes to their robust possession and exercise of these capacities, as it means even agents who aren't ideal can succeed in acting in ways that elude criticism from the standpoint of reason. But there are plainly agents whose possession and exercise of these capacities is even more robust.

Imagine agents who robustly have and exercise the capacity to realize their desires, *no matter what their content*, and to know what the world is like, *no matter what it is like*. Such agents are even more ideal than the agents we have described thus far because their success in eluding criticism isn't tied to their having the particular desires that they have, or to the worlds' being the particular way that it is. Since such agents are in this way *invulnerable to contingency*, it follows that their possession and exercise of the capacities for knowledge-acquisition and desire-realization is *maximally robust*. But if this is what it is for an ideal agents' possession and exercise of their knowledge-acquisition and desire-realization capacities to be maximally robust, then it follows immediately that there must be more to criticism from the standpoint of reason than we have seen so far. Moreover, as we will see, the more that there is entails that harming someone is criticizable from the standpoint of reason: in other words, there is a reason not to harm.

5. What is the relationship between the standpoints of morality and reason?

Ideal agents—that is, those agents who are immune to criticism from the standpoint of reason—are those who robustly have and exercise the capacity to realize their desires, no matter what their content, and to know what the world is like, no matter what it is like. Let's call this the modal interpretation of what it is to be ideal. If ideal agents, understood according to the modal interpretation, are so much as possible then there must be more to criticism from the standpoint of reason than we have seen so far. If there weren't, then among the intrinsic desires ideal agents could have and realize are intrinsic desires whose realization is inconsistent with their being ideal. In other words, ideal agents would be impossible.

In order to see that this is so consider the predicament of an otherwise ideal rational agent who has an intrinsic desire to (say) believe that grass is red rather than green. If there weren't more to criticism from the standpoint of reason than we have seen so far then such an agent would not be subject to criticism from the standpoint of reason for having and realizing such a desire. But in possible worlds like ours in which grass is green, agents who have such an intrinsic desire could not avoid criticism from the standpoint of reason, as they could not simultaneously exercise their capacity to realize their desires, no matter what the content, and their capacity to know what the world is like, no matter what it is like. Either their intrinsic desire to believe that grass is red would go unrealized, albeit they could exercise their capacity to know that grass is green, or they

could exercise the capacity to realize their intrinsic desire to believe that grass is red, but would thereby be precluded from knowing the color of grass.

But what more could there be to criticism from the standpoint of reason? My conjecture is that, in virtue of being ideal, agents must have to care about certain things, namely, those things which are such that, if they were to bring them about, then the potential for the kind of conflict just described in the exercise of their capacities for desire-realization and knowledge-acquisition would disappear. The obvious intrinsic desire for an ideal agent to have in order to resolve such a conflict is a generally dominant intrinsic desire not to interfere with their exercise of their capacity to know what the world is like, no matter what it is like, either now or in the future. For if ideal agents have such an intrinsic desire, then even if they do intrinsically desire to believe that grass is red, they would realize the former intrinsic desire rather than the latter, as the exercise of their capacity reliably to realize desires, no matter what their content, is sensitive to the *strengths* of the different desires that they have.

Nor is this the only intrinsic desire that ideal agents would have to have. Given that agents exist over time, a similar potential for conflict arises between the exercise of the capacity for desire-realization in the present and the exercise of the capacity for desire-realization in the present and the future. Since the satisfaction of present intrinsic desires could interfere with the satisfaction of present and future intrinsic desires—imagine someone whose strongest present intrinsic desire is not to act in a globally instrumentally rational way in either the present or the future—the same line of reasoning suggests that ideal agents—these are agents who robustly possess and exercise their capacities at each moment that they exist—would also have to have a generally dominant intrinsic desire not to interfere in the present with the exercise of their capacity to realize their intrinsic desires in the present or the future, on condition that the realization of their present and future intrinsic desires wouldn't lead them to interfere with the exercise of the knowledge-acquisition or desire-realization capacities of themselves in the further future.

(From here-on I will take the latter condition as read, but before proceeding it is worth pausing to clarify the content of the desires already postulated, especially the latter, given the condition. Note that intrinsic desire not to interfere in the present with the exercise of one's capacity to know what the world is like in the present or future, and the intrinsic desire not to interfere with the exercise of one's capacity to realize one's intrinsic desires in the present or the future, on condition that the realization of one's future intrinsic desires wouldn't themselves lead to interference with the exercise of one's knowledge-acquisition or desire-realization capacities, are not the same as a present intrinsic desire that non-interference with the exercise of one's desire-realization and knowledge acquisition capacities in the present and future be maximized. Someone with the latter desire would be indifferent between presently interfering in the future, so ensuring that there is no interference of a similar kind in the future, and not interfering in the present, thereby allowing that future interference. But someone with the former intrinsic desire would not be indifferent. Moreover, if one's future self possesses the capacity for self-control, then it seems that one shouldn't be indifferent.

One of the ways in which one could presently interfere with the exercise of the desire-realization capacities of one's future self, when one's future self possesses the capacity for self-control, is by pre-empting the opportunity of one's future self to exercise self-control. Not interfering with the exercise of the desire-realization capacities of one's future self, when one's future self possesses the capacity for self-control, must therefore amount to *leaving it up to one's future self whether or not there will be interference*. If this is right then someone with a generally dominant intrinsic

desire not to interfere in the present with the exercise of their capacity to realize their intrinsic desires in the future, on condition that the realization of their future intrinsic desires wouldn't lead them to interfere with the exercise of the knowledge-acquisition or desire-realization capacities of themselves in the further future, may well have to tolerate a certain amount of interference by their later self as the acceptable cost of their not interfering with their later self.)

There are other intrinsic desires that ideal agents would have to have as well. According to the modal interpretation of what it is to be ideal, ideal agents have the capacity to realize their desires, no matter what their content, and to know what the world is like, no matter what it is like, and they exercise these capacities at each available opportunity throughout the duration of their existence. But given that the development and maintenance of these capacities is to some extent under the control of agents themselves, we have to ask whether we can coherently suppose that an ideal agent would sit idly by and fail to develop their capacities for knowledge-acquisition and desire-realization, if developing them was an option, and if they already had these capacities, whether they would sit idly by and watch themselves lose their capacities for knowledge-acquisition or desire-realization, if maintaining them was an option. The answer would seem to be that we cannot imagine this. Ideal agents would do what they can to help develop and maintain their capacities. Ideal agents would therefore have to have an additional generally dominant intrinsic desire to do what they can to ensure that they develop and maintain their capacities for knowledge-acquisition and desire-realization, both in the present and in the future.

If ideal agents have all of these generally dominant intrinsic desires, and if they also have and exercise the capacity to know what the world is like, no matter what it is like, and realize their desires, no matter what their content, then note that there is no longer the possibility of a conflict in the exercise of their capacities of the kinds we saw earlier. There is, of course, the possibility of a conflict between the different generally dominant intrinsic desires that ideal agents have to have. Imagine a situation in which the only way in which an agent could maintain their capacities for knowledge-acquisition and desire-realization is by interfering with the exercise of some knowledge-acquisition or desire-realization capacity that they have. But the governing idea of an agent who robustly possesses and exercises the capacity to know what the world is like, no matter what it is like, and to realize their desires, no matter what their content, will itself give us such guidance as we can have about how such conflicts are to be adjudicated: that is, our conception of an ideal agent will tell us which of these generally dominant intrinsic desires should dominate which in the circumstances.

So far we have seen that ideal agents must have a generally dominant intrinsic desire not to interfere with the exercise of their knowledge acquisition capacities both in the present and the future, a generally dominant intrinsic desire not to interfere with their exercise of their desire-realization capacities in the present and the future, and a generally dominant intrinsic desire to do what they can to help ensure that they develop and maintain these capacities. For short, let's call these the desires to help but not interfere with their rational capacities. We have seen that they must have these desires because only so will their possession and exercise of these capacities be robust. They could just so happen to have desires whose realization doesn't interfere with their possession and exercise of their rational capacities, but this would make them vulnerable to the contingency that that they happen to have the desires that they have, or the world happens to be the way that it is.

Having said that, note that the desires to help but not interfere, at least as we've understood them thus far, are restricted to *the agent's own* present and future exercise of their rational capacities. The question is whether ideal agents' desires to help but not interfere would be restricted to themselves in this way, or whether they would extend their concern to other agents as well. Here we note some striking symmetries. Think again about the way in which an agent's later self's possession and exercise of their rational capacities is vulnerable to what their present self desires to do. Even if their present self doesn't desire to interfere, and even if the later self doesn't need help, the later self's possession and exercise of their rational capacities is less robust than it would have been if the present self had had generally dominant desires to help but not interfere with their later self. Since the modal conception of what it is for an agent to be ideal tells us that an ideal agent's possession and exercise of their rational capacities is maximally robust, we are therefore forced to conclude that ideal agents' present selves must have generally dominant desires to help but not interfere with the possession and exercise of the rational capacities of their later selves.

The striking symmetry is that *each agent stands in much the same relation to other agents* as their later self stands in to their present self Among the desires that ideal agents can have and realize are *social* desires, that is, desires whose realization requires that they interact with other agents. Now try to imagine an ideal agent, as we have so far characterized ideal agents, surrounded by other agents—let's call these the ideal agent's *world-mates*—and ask what needs to be the case for the ideal agent's possession and exercise of his rational capacities to be maximally robust in such a world. Though we can certainly imagine that his world-mates don't interfere with him, and that he doesn't need their help, his possession and exercise of his rational capacities would be modally fragile. His possession and exercise of his rational capacities would be more robust if his world-mates happened to have generally dominant desires to help and not interfere with him, as in that case he would still possess and exercise his rational capacities to a high degree in those nearby possible worlds in which his world-mates do desire to interfere with him, and he can't do all that's required to develop and maintain his rational capacities by himself, as his world-mates would do what's required. But his possession and exercise of his rational capacities would still be modally fragile because there are also nearby possible worlds in which his world-mates don't have a generally dominant desire to help him and do have a generally dominant desire to interfere with him, and in which their capacities for the realization of their own desires leave him defenseless.

In order to imagine an ideal agent surrounded by other agents—that is, an agent who in nearby possible worlds still manages to a very high degree to know what the world is like, no matter what it is like, and to realize his desires in that world, no matter what their content—we must therefore suppose that helping but not interfering are themselves actions that his world-mates perform in a far greater range of nearby possible worlds. But can we really imagine that that is possible? The answer is that that we can. If ideal agents aren't just concerned with their own present and future possession and exercise of their rational capacities, but extend their concern to all agents, both in the present and the future, and if we begin with the supposition that in those worlds in which our imaginary agent is an ideal agent with social desires, his world-mates are also ideal agents, then he will still possess and exercise his rational capacities to a very high degree in a great many of the nearby possible worlds in which his world-mates don't have a generally dominant desire to help him, but do have a generally dominant desire to interfere with him, and in which their capacities for desire-realization leave him defenseless. This is because, in

a great many of these nearby possible worlds, his world-mates know that their ideal counterparts would desire that they help but don't interfere with him, and they exercise the self-control required to get themselves to do just that despite the fact that they aren't antecedently inclined to do so.

As advertised, the situation is thus symmetrical. Because the modal conception of what it is for an agent to be ideal tells us that ideal agents' possession and exercise of their rational capacities is *maximally* robust, we are forced to conclude that each ideal agent must have generally dominant intrinsic desires to help but not interfere with the possession and exercise of not only their own rational capacities in the present and in the future, but *any agent's* rational capacities in the present and in the future. Indeed, more than this, we are forced to conclude that in those possible worlds in which ideal agents have social desires, they are themselves surrounded by other ideal agents. To put the crucial point in the form of an aphorism, this is because *each of us is better for being in the presence of the better selves of others*. Or, to put the same point in the form of a bumper sticker: *Better Together*. Or, to put it in the form of a scholarly observation: *Kant was right that understanding what it is to act in ways that elude criticism from the standpoint of reason leads us to postulate a Kingdom of Ends*.

6. Conclusion

Let's sum up. We began by asking whether there are any normative standpoints that are thoroughly global and non-idiosyncratic. We noticed that there are two such standpoints, the standpoint of reason and the standpoint of morality, and this made us wonder what the relationship is between these two standpoints.

We have seen that the agents who elude all criticism from the standpoint of reason are those who robustly possess and exercise maximal capacities for knowledge-acquisition and desire-realization, and that such agents have generally dominant intrinsic desires not to interfere with any agents' exercise of their knowledge-acquisition or desire-realization capacities, and generally dominant intrinsic desires to help all agents acquire and maintain such capacities. Since helping but not interfering are also what's plausibly required to elude criticism from the standpoint of morality, the upshot is that the two standpoints are intimately related. The standpoint of morality is grounded in the standpoint of reason. The two moral principles Help, but Do Not Interfere, are therefore grounded in two generally dominant reasons for action, reasons to help but not interfere. Or, to put the same point in terms of the moral prohibition on harming, the moral principle Do No Harm splits into two principles that correspond to the two ways in which we can harm someone, by interfering with them and by failing to help them.

What is the commonsense content of these two moral principles? Not interfering with the exercise of any agents' knowledge-acquisition and desire-realization capacities amounts to leaving everyone free to make up their own minds about the nature of the world in which they live and about how they will lead their own lives in that world. Helping agents acquire and maintain these capacities amounts to ensuring that everyone has the wherewithal to make up their own minds about the nature of the world in which they live, and the wherewithal to lead lives of their own choosing in that world. More materially, these two moral principles suggest that, if we are to elude criticism from the standpoint of reason, it is incumbent upon all of us to contribute to a social system in which basic education, healthcare, and equality of opportunity are enjoyed by everyone, and which enjoins each of us to leave others free to lead lives on

whatever terms they please, so long as their terms don't impinge on the freedom of others to lead their lives on whatever terms they please.

To the extent that these sound like the basic tenets of liberal morality, the upshot is that liberal morality is itself mandated from the standpoint of reason. Moreover, as we have seen, nothing beyond an understanding of ourselves as rational agents is required to see why this is so.

REFERENCES

Hume, David 1740: *A Treatise of Human Nature* (Oxford: Clarendon Press, 1968).

Kant, Immanuel 1786: *Groundwork of the Metaphysics of Morals* (London. Hutchinson and Company, 1948).

Mackie, J. 1977: *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin).