Reason, Sentiment, and Categorical Imperatives

Morality is based in reason. That slogan is (and ought to be) associated with Kant’s ethics. But the slogan’s simplicity should not obscure that it represents a number of distinct and complex claims. Among them is the following: According to common sense, morality contains categorical imperatives; and such imperatives can be grounded only in reason. Here I defend this claim against a sentimentalist opponent.

After summarizing a sentimentalist moral theory, I argue that it is incompatible with the notion that there are absolutely necessary and universal obligations, that is, categorical imperatives. Such imperatives cannot be grounded in feelings. I then contend that the account of obligation suggested by sentimentalism clashes with ordinary moral thinking in ways that a Kantian account does not. In particular, sentimentalism is unable to accommodate the widespread judgment that, in many circumstances, members of a group have an obligation not to manipulate, exploit, or coerce outsiders, even if the members believe that doing so will benefit their group. I resist the sentimentalist suggestion that we forfeit nothing significant if, unable to assert that the members of the group are obligated to treat outsiders decently, we simply designate the members’ behavior as vicious or odious when they fail to treat them decently. Finally, I underscore the plausibility of believing in categorical imperatives. I suggest that it is no more odd to believe in them than it is to hold “the hypothetical imperative” to be valid. It is even reasonable to think that sentimentalism (or at least the version of it I consider) is itself committed to there being a categorical imperative.

The main claim I defend is not transparent. It is not obvious what a categorical imperative
is supposed to be, let alone what it would mean for one to be grounded in reason. But the meaning of the claim comes to light if we examine it in relation to several other claims underlying the Kantian slogan that morality is based in reason.

**Reason and Morality according to Kant**

To begin, Kant contends that philosophical argument rationally compels all of us to hold the Categorical Imperative to be valid. In one well-known version, the Formula of Humanity, the Categorical Imperative commands: “So act that you treat humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means” (Kant, 1996b, p. 80, italics omitted). To hold this principle to be valid is to maintain that it is unconditionally binding on all rational agents; we always ought, all things considered, to act in accordance with it, no matter what we might desire to do. In the third section of the *Groundwork of the Metaphysics of Morals*, Kant unfurls a complex philosophical argument, which he calls a “deduction,” in an attempt to demonstrate the Categorical Imperative’s validity. If the attempt succeeds, then, in Kant’s view, he establishes that this imperative is the supreme principle of morality.

But Kant soon abandons the attempt to construct a deduction of the Categorical Imperative. In the *Critique of Practical Reason*, published three years after the *Groundwork*, he asserts that “the objective reality” of the Categorical Imperative “cannot be proved by any deduction” (Kant, 1996a, pp. 177-78). For present purposes, let us call a philosophical argument one that rests (or purports to rest) on premises that are demonstrably irrational to deny since, for example, it can be shown that denying them involves self-contradiction. Kant here seems to
claim that no philosophical argument can establish the Categorical Imperative’s validity.

Nevertheless, he says, the Categorical Imperative (or the moral law) “is given, as it were, as a fact of pure reason of which we are a priori conscious and which is apodictically certain” (Kant, 1996a, p. 177); the Categorical Imperative is “firmly established of itself” (Kant, 1996a, p. 178). Precisely what Kant means here is debatable, but I think he implies at least the following. Since we have a certain capacity, namely reason, the Categorical Imperative is binding on us. Each rational agent’s own reason demands that he abide by the Categorical Imperative. Moreover, each rational agent is able to realize that he is bound by this principle. So the Kantian slogan that morality is based in reason also contains the claim that it is partly constitutive of being a rational agent to be bound by the Categorical Imperative and to be able to realize that one is so bound.

In a third sense, the dictum that morality is based in reason amounts to the claim that for all agents all the time, it is irrational to act contrary to the Categorical Imperative. Immoral action is irrational action. Our own reason sets forth this imperative, so it is contrary to reason to disobey it. For Kant the irrationality of immoral conduct is a function of its clashing with a valid demand of reason. If the Categorical Imperative was not valid, then in Kant’s view there would not necessarily be anything at all irrational in someone’s pursuing his goal of getting rich by making a false promise or always refusing to help others in need.

A fourth Kantian claim that lies beneath the slogan that morality derives from reason takes shape against the background of a terminological distinction. Kant and Kantians use “categorical imperative” in several different ways. I employ “Categorical Imperative” (capitalized) to refer to the imperative that Kant takes to be the supreme principle of morality. I use the term “categorical imperative” (lower case) in a more general sense, namely to refer to any
principle that has two characteristics. The first is that of absolute necessity. A principle is
absolutely necessary just in case all agents within its scope are, all things considered, obligated to
do what it specifies, and, moreover, their obligation is unconditional in that it does not depend on
the agents’ having any particular desires or ends.¹ An absolutely necessary principle can be
“overridden” neither by any other principle, nor by any inclination or need, no matter how
pressing. A categorical imperative’s second characteristic is universality, that is, a scope that
extends to all rational agents. According to Kant, of course, the Categorical Imperative counts as
a categorical imperative. (That’s another way of saying that the Categorical Imperative is valid.)
But we need to keep in mind that a categorical imperative is any principle, invoked by Kant or
not, that has these characteristics. A fourth claim implicit in the slogan that morality is grounded
in reason is the following. According to reflective common sense, morality contains categorical
imperatives. But philosophical inquiry reveals that the only way there can be categorical
imperatives is if they are demands placed on rational agents by the agents’ own reason.
Categorical imperatives cannot exist if morality is based entirely on God’s will, custom,
sentiment and so forth. Kant defends this claim (at least implicitly) in section 2 of the

Groundwork of the Metaphysics of Morals.

A fifth Kantian claim implicit in the slogan that morality is based in reason invokes
Kant’s concept of moral worth. Kant holds that an action has moral worth if and only if it is done
from duty. On my interpretation, Kant maintains that an agent acts from duty just in case: her
motive for acting stems from the notion that some categorical imperative requires the action; this
notion itself provides sufficient motive for her acting; she acts against the background of
conscientious reflection; and she does her best to realize her action’s end (Kerstein, 2002, p.

-4-
Actions done from duty involve reason in two main ways. First, a product of reasoning, namely the idea that some categorical imperative requires the action, not only motivates the action but itself provides a sufficient motive for it. When an agent does something from duty, the ultimate source of her motivation to do it can be simply the thought that doing it is required by some unconditionally and universally binding principle. In Kant’s terminology, pure reason can be practical (Kant, 1996a, p. 165). Actions done from duty invoke reason in a second way in that they involve (what I call) “conscientious reflection.” Since, when someone acts from duty, he is interested in doing what is morally required, he must take (or have taken) an active interest in finding out just what that is. The person need not delve into casuistry before every action, but he needs to act against the background of reflection on the moral status of what he does. Such reflection involves reasoning—for example, determining whether a particular sort of action is forbidden by some principle. For Kant morally worthy action necessarily involves the use of reason, both as a tool for discovering that the action is morally required and as a motive for doing it. This is the fifth claim.

These five claims do not exhaust the ways in which, according to Kant, morality is based in reason. But when someone defends the Kantian slogan, chances are she is championing at least one of them. Listed together, the claims are:

1. Philosophical argument rationally compels all of us to hold that the Categorical Imperative is valid.
2. It is partly constitutive of being a rational agent to be bound by the Categorical Imperative and to be able to realize that one is so bound.
3. It is always irrational for an agent to act contrary to the Categorical Imperative.
(4) According to everyday moral thinking, there are categorical imperatives; moreover, categorical imperatives can only be demands placed on rational agents by the agents’ own reason.

(5) Morally worthy action necessarily involves the use of reason, both as a means taken to discover that the action is required and as a motive for doing it.

A few points regarding the relations between these claims warrant attention. The truth of the second claim does not seem to depend on that of the first. No philosophical argument, let us suppose, is so good that it would be irrational for anyone to deny that it demonstrated the Categorical Imperative to be valid. It might nevertheless be part of what it means to be a rational agent to be bound by this imperative and to be able to realize that one is. Second, for Kant the third claim derives any force it has from the first or the second. If it is true neither that philosophical argument rationally compels all of us to hold the Categorical Imperative to be valid, nor that it is constitutive of being a rational agent to be bound by this imperative, then from Kant’s perspective there is no basis for claiming that it is always irrational for an agent to act contrary to the Categorical Imperative. So for him the third claim is derivative. A final point to notice about the relations among the five claims is that the last two might be true even if the first three are all false. Neither (4) nor (5) invokes the Categorical Imperative. To focus on (4), suppose that part of the content of morality is constituted by universal and unconditional imperatives placed on agents by their own reason. If we suppose this, it follows neither that among these imperatives agents are able to find, nor that are rationally compelled to find, the Categorical Imperative—nor, of course, does it follow that acting contrary to the Categorical Imperative is irrational. Some of Kant’s claims regarding morality’s being based in reason are
independent of his advocacy of the specific principles that count as formulas of the Categorical Imperative, that is, principles such as the Formula of Humanity.

Defending Kant’s Claims

Perhaps the most dramatic claim of the five and the one that a Kantian would be most eager to defend is the first, namely that philosophical argument rationally compels us to hold the Categorical Imperative to be valid. I would like to defend it. But the deduction of the Categorical Imperative that Kant undertakes in *Groundwork* III is notoriously difficult to interpret, and on no interpretation I am aware of does the argument succeed. For example, on one reading the argument unfolds (in very broad outline) as follows: ‘Since we take ourselves to be rational, we must assume that we are free, that is, capable of acting spontaneously. But spontaneous action must be governed by some law. And this law could only be the Categorical Imperative. Therefore, we must assume that we are bound by the Categorical Imperative.’ Each step of this argument is controversial. But even if we accept the first and second steps, it remains unclear why we should embrace the third. What argument establishes that the Categorical Imperative is the only law capable of governing a free will? I simply do not know how to make Kant’s deduction of the Categorical Imperative work. Of course, even if no one else at present does either, it does not follow that no one will make it work.

Some contemporary philosophers have defended claims that bear a family resemblance to (1). Christine Korsgaard, for example, tries to demonstrate that unless we take humanity to be valuable, we condemn ourselves to complete practical skepticism, that is, to the view that we have no reason to do anything at all (Korsgaard, 1996, p. 122). In other words, if we assume that
we have reasons for our actions, then we must value all persons. If Korsgaard’s project is successful, then a philosophical argument rationally compels those of us who hold that we have reasons for what we do (i.e., almost all of us) to embrace a morally significant conclusion. An argument that did this would perhaps not be as dramatic as one that rationally compelled us to hold the Categorical Imperative to be valid, but it would be striking nonetheless. Yet, as I explain elsewhere (Kerstein, 2001), I do not believe that Korsgaard’s argument succeeds. I am aware of no convincing contemporary effort to establish a claim akin to (1). Nevertheless, it is, I think, premature for us to give up hope that a proof will emerge.

But even if we do, we are not forced to give up the slogan that morality is based in reason. As is now evident, this slogan incorporates several claims that do not depend on (1). Elsewhere (Kerstein, 2002) I have endorsed some aspects of (5). I think it is plausible to claim that all actions from duty involve reason, and that all such actions have moral worth. But my main aim here is to defend claim (4), namely that according to everyday moral thinking there are categorical imperatives, and that these imperatives can only be demands placed on rational agents by their own reason. In particular, I want to defend this claim against sentimentalism—a philosophical view that vehemently rejects the idea that morality is based in reason.

The claim at issue has two components. The first is relatively straightforward, namely that, according to reflective common sense, there is at least one principle of conduct that is absolutely necessary and universal binding. The second component is a bit more complex. What would it mean for categorical imperatives to be demands placed on rational agents by their own reason? To answer this question thoroughly from the perspective of Kant’s ethics would require a long discussion of his theory of agency and his concept of autonomy. Since this is not the place
for such a discussion, it makes sense to appeal to aspects of claims (1) and (2) above and to maintain the following. A demand is placed on rational agents by their own reason only if either we are all rationally compelled by philosophical argument to abide by it, or it is partly constitutive of being a rational agent to be bound by it. According to this latter possibility, any failure to honor the demand constitutes a form of irrationality, unless it is sanctioned by some overriding demand that is itself based in reason. The burden of defending the second component is to show that agents’ own reason is the only viable candidate for the source of a categorical imperative. In other words, the only way that a principle can have the universality and absolute necessity distinctive of a categorical imperative is for it to be a demand of reason. It will not have it simply by virtue of being commanded by God, or being such that adherence to it will maximize the general welfare, and so forth.

This is a significant burden, and I do not assume it all here. But I do attempt to support the second component of (4). I try to show that a version of sentimentalism recently developed by Simon Blackburn, my opponent in this debate, is incompatible with the notion that there are categorical imperatives. I then argue that this incompatibility is not a mere philosophical curiosity, but brings sentimentalism into conflict with ordinary moral thinking. I thus offer some support for the first component of (4), namely the claim that in our everyday moral reflection we endorse the notion that there are categorical imperatives. Sentimentalism pays a high price for denying that morality is grounded in reason.

In what follows, I take Blackburn’s sentimentalism, which is inspired by the theories of David Hume and Adam Smith, as representative of sentimentalism as a whole. For ease of expression, I will refer to his theory simply as sentimentalism. Of course, there are many
different versions of sentimentalism, and some might be less vulnerable to the Kantian criticisms presented here than Blackburn’s. But I am skeptical of whether any sentimentalist account will be immune to them.

Sentimentalism and Obligation

Blackburn offers a succinct summary of sentimentalism’s core:

First we love one or another quality in people when we come across it, possibly because we have been educated to do so. Then we take up the common point of view which turns love to esteem, assessing a trait of character as admirable or the reverse. Third, we can become aware that this is a trait that we ourselves exhibit, or do not. And fourth, when we do so we are moved to a self-satisfaction and pride, or unease and shame, corresponding to our original assessment, and imagining this assessment made of us by others. This is a kind of internal vibration in sympathy with the imagined sentiments of others. (Blackburn, 1998, p. 203)

Blackburn cites with apparent approval Hume’s notion that we love character traits that render those who possess them useful or agreeable to themselves or to others. So suppose a person loves the character trait of being just to strangers. (Someone with this trait will have a disposition, for example, to refrain from cheating people he does not know.) The person then takes the “common point of view” and ends up esteeming this trait as admirable. The common point of view does not amount to the perspective of an impartial spectator, Blackburn suggests. To take this view is not to take into account the interests of all persons, but rather to abstract from one’s own position and
consider the impact that a trait of character has on “a fairly immediate circle,” including the person who possesses it and, for example, his family and friends (Blackburn, 1998, p. 210). Moreover, to take the common point of view is to embrace a certain kind of civility to others. It is to refuse to “rest content with relations with outsiders that we cannot see ourselves justifying to them” (Blackburn, 1998, pp. 210-211). So presumably in the process of discerning whether the character trait in question is admirable, the person would consider whether he, in acting in accordance with it, could see himself justifying his action to those outside of his circle. A person who had a disposition to be just to strangers would exhibit civility, it seems. Proceeding through Blackburn’s account, the person then becomes aware of whether she has the character trait of being just to strangers. Let us suppose that she finds that she does not. Imagining others’ disapproval at not having this trait, she experiences the displeasing sentiments of unease and shame.

Blackburn’s summary leaves unclear how sentimentalist theory makes room for requirements to perform or refrain from performing certain actions. How do moral obligations stem from the process he describes? To my knowledge, Blackburn does not answer this question. But he would, I gather, embrace an answer suggested by Hume: “[W]hen any action or quality of the mind, pleases us after a certain manner, we say it is virtuous; and when the neglect, or non-performance of it, displeases us after a like manner, we say that we lie under an obligation to perform it” (Hume, 1978, p. 517, italics omitted). So it is the person’s displeasing sentiments, ones such as unease or shame, that form the basis of her obligation to acquire the character trait of being just to strangers, or at least to act in a way that a person with this trait would act.

This last point is crucial to the issue of whether sentimentalism coheres with the idea that
there are categorical imperatives. On this account, the basis for an agent’s obligation to do something is a displeasing sentiment she has when, after taking the “common point of view,” she contemplates her not doing it or, perhaps, her not possessing the character of someone who would do it. (For the sake of simplicity, I omit mentioning this latter possibility below.) If an agent does not have this sentiment, then she has no obligation. Of course, if an agent has no obligation to perform a certain action, then a principle commanding that action does not count as a categorical imperative. For it belongs to the concept of a categorical imperative that everyone within its scope is obligated to do what it enjoins. So in order for sentimentalism to ground a particular categorical imperative, each and every person, after taking the common point of view and so forth, must have a displeasing sentiment towards not doing what the imperative commands.

Yet is it plausible to think that everyone would? Blackburn recognizes that there are people who “confine their concerns to their immediate tribe or group or class or gender,” and for whom others “don’t count as much or just don’t count at all.” These people might even gain in reputation and honor “by callous or fraudulent dealings” with others outside of their circle. Blackburn calls these people “foreign office knaves,” evoking the image of a colonial administrator exploiting a native population (Blackburn, 1998, p. 211).³ This term is acceptable, as long as we keep in mind that the vast majority of foreign office knaves are found outside of any foreign office. It is naive to deny that they exist in virtually all walks of life, and in an extremely wide variety of cultures (both historical and extant). Limiting ourselves to a slice of the here and now, the Hutus and the Tutsis each have their share of foreign office knaves, and so do the Serbs and Croats, the Palestinians and Israelis, and so forth.
With foreign office knaves in view, take the following candidate for a categorical imperative: “Do not treat strangers unjustly.” (Let us assume that unjust treatment includes fraud, exploitation, and so forth.) Obviously, the knaves might have no displeasing sentiment towards failing to treat strangers justly. On the sentimentalist account, such a sentiment would emerge from taking the common point of view. But a foreign office knave might not take this point of view. In particular, he might lack the impetus to think about whether he could see himself justifying his actions to outsiders. What the knave cares about is his own welfare and the welfare of his people. A foreign office knave who does not take the common point of view and who thus has no displeasing sentiment towards treating strangers outside of his group unjustly would, according to sentimentalism, (at least sometimes) have no obligation to act in accordance with the principle “Do not treat strangers unjustly.” Sentimentalism would therefore not allow this principle to stand as a categorical imperative.

Of course, that does not entail that it would allow no principle to stand as such. But it does cast serious doubt on the possibility. Take any action. At least one person would fail to have a displeasing sentiment at the prospect of not performing it, it seems reasonable to conclude; and such a person would not be obligated to perform the action. So an imperative prescribing the action’s performance would not count as a categorical imperative.

As I mentioned, Blackburn does not explain how moral obligations stem from moral sentiments. So perhaps he would reject the account suggested here, according to which an agent is obligated to do P just in case, after he takes the common point of view, he has a displeasing sentiment towards not doing P. Might a different account of sentimentalist obligation support categorical imperatives? Perhaps, but I do not know what that account would be. Someone might
try to render sentimentalist obligation less subjective and contingent by claiming that an agent is obligated to do P just in case, *if he took the common point of view, then he would have a displeasing sentiment towards not doing P.* On this account, obligation would stem not from an agent’s actually having a certain sentiment after taking the common point of view, but rather from the fact that he would have this sentiment if he did take the common point of view. But I do not think this hypothetical account helps very much.

Consider again a foreign office knave. In taking the common point of view, he would be asking himself (roughly) whether an individual’s exploiting foreigners would be beneficial to that individual’s circle. And the answer to that question might, of course, be yes. The knave would also be asking himself whether, in accordance with a notion of civility, he (the knave) could see himself justifying his action to foreigners he was exploiting. Now this is a difficult question to answer, since Blackburn leaves the notion of civility very vague. But the knave might believe that he could fulfill it by saying something like this to those he exploited: ‘What I have done to you, I consider to be morally unobjectionable. But I would also consider it morally unobjectionable for you to do the same to me (though I would not be happy about it). It is morally permissible for each of us to pursue the good of our own circle.’ So even if he takes the common point of view, the knave might have a pleasing sentiment towards exploiting foreigners, and so not be obligated to refrain from doing so.

One might object that the knave has not here really taken the common point of view, since he has not really embraced the notion of civility. But is it clear that he has not? The knave does at least see himself to be offering a minimal justification; for he underscores that he is not making an exception of himself: the same standard applies to him as does to those he exploits.
(He is appealing to something like what I later call the principle of reasons universalism.) To
dismiss the knave’s effort as unacceptable, we would need a more thorough account of the
distinction between appropriately and inappropriately seeing oneself as able to justify one’s
actions to outsiders. In any case this objection has little impact. For the sake of argument let us
grant that if the knave properly took the common point of view, then he would find that in
exploiting others, he would fall short of exhibiting civility. The problem is that for him this
realization might not be accompanied by a disagreeable sentiment. Again, the foreign office
knave just does not care about anyone outside of his circle. So why think that he would have such
a sentiment? Why assume that taking the common point of view would work emotional magic?
If, as seems plausible to assume, it would not, then he would not be obligated to refrain from
exploiting foreigners. Once again, we find a version of sentimentalism unable to support the
principle ‘Do not treat strangers unjustly” as a categorical imperative.

These considerations do not prove that no form of sentimentalism could support any
categorical imperative. But they do, I believe, cast significant doubt on the possibility of
grounding universal, unconditional obligation in moral sentiments. Sentimentalism does not
really threaten Kant’s claim that categorical imperatives can only be demands placed on rational
agents by the agents’ own reason.

No Categorical Imperatives, No Problem?

At the tip of the sentimentalist’s tongue might be the following rejoinder. So what if
sentimentalism is incompatible with the notion that there are categorical imperatives? What does
it matter? It matters, I think, because this incompatibility has implications unacceptable to
reflective moral common sense. Showing that it does will lend some support to Kant’s contention that in our moral thinking we are committed to there being categorical imperatives.

First, sentimentalism must deny that people have obligations when many of us are convinced they do. This point emerges clearly from reflection on a thought experiment suggested by Hume. He invokes the image of “a species of creatures intermingled with men, which, though rational, were possessed of such inferior strength, both of body and mind, that they were incapable of all resistance, and could never, upon the highest provocation, make us feel the effects of their resentment.” (Hume, 1975, p. 190). We, that is, the human beings intermingled with these creatures, would have no obligation to treat them justly, says Hume. For justice is grounded solely in public utility, which in this case is our utility, and the creatures’ weakness renders it useless for us to treat them justly. Nevertheless, Hume suggests that we are “bound by the laws of humanity to give gentle usage to these creatures” (Hume, 1975, p. 190). Apparently, we are so bound ultimately because treating them well would give us pleasing sentiments (or abusing them would give us displeasing ones). Hume calls our “compassion and kindness the only check, by which [these weak rational creatures] curb our lawless will” (Hume, 1975, pp. 190-91). But now suppose that these weak rational creatures were intermingled with a group of foreign office knaves. Devoid of displeasing sentiments towards exploiting these creatures, the knaves would, on Blackburn’s (and presumably Hume’s) sentimentalist account, have no obligation to treat them decently: to refrain from exploiting, enslaving, or even torturing them. To many of us, this conclusion seems shocking. But a belief that morality embodies categorical imperatives puts us in position to resist it. For it is open to us to maintain that the foreign office knaves are obligated to treat the weak rational creatures decently, since not to do so is to violate a
universally and unconditionally binding principle—for example the Kantian imperative not to treat rational agents merely as means.

Second, holding that moral obligations are a matter of sentiment rather than categorical imperatives has implausible implications regarding temporal features of obligations. Imagine you have been raised to your late teens in an insular, protected, and somewhat clannish social group. Prevalent in it is an antipathy towards an ethnic minority whose members are repeatedly represented as lazy, untrustworthy, and vicious. Your concerns are limited almost exclusively to your own happiness and to that of your group. The prospect of exploiting members of this minority gives you no displeasing sentiment whatsoever. You do not ask yourself (and do not care) whether you could justify to them treating them in one way or another. You do not take the common point of view. On Blackburn’s sentimentalist account, you have at this point no obligation to refrain from abusing the minority. (That is disconcerting, to say the least, but it is not the issue here.) Now suppose that, much to your chagrin, you and a member of the despised minority must cooperate in the workplace. At the beginning things are tense, but slowly you warm to him and your attitude towards his ethnic group becomes less hostile and dismissive. You do not mention your change of attitude in your own group until one of your friends hatches a plan to cheat a member of the minority out of her paycheck. Almost to your surprise, the prospect of carrying out the plan is very disagreeable to you. You tell your friends that you will not participate, and that it would be better if they left the woman alone. The sentimentalist might say that with your change in attitude, you have come to have an obligation not to exploit the minority. You did not have the obligation before, but you do now. But how would you think of the situation? Would you not believe rather that before your work experience, you had an
obligation not to mistreat the minority, but that you were, regrettably, unaware that you did? That
the obligation itself suddenly appeared with the sentiment seems an unlikely thing for you to
hold, though it would be quite natural for you to believe that the sentiment helped you become
cognizant of the obligation.

If it is odd to think of a person’s acquiring an obligation simply through coming to have a
certain sentiment, it is perhaps even stranger to think of him as forfeiting it if his sentiments
change. Suppose that members of your own group were so persistent and clever in promoting
loyalty to them and condemning the minority that your sentiments underwent yet another turn.
You found yourself once again unwilling to take the common point of view and devoid of
displeasure at the prospect of mistreating the minority. Your obligation not to mistreat the
minority would thereby dissolve, according to sentimentalism. But obligations do not seem to
pop in and out of existence like that, or so many of us believe. A moral theory that can embrace
categorical imperatives can respect this belief. For categorical imperatives would be
unconditionally binding.

Is Obligation Overrated?

Here a sentimentalist might object that we have been making entirely too much of the
notion of obligation. He might point out, for example, that though on his account we cannot say
that the foreign office knaves have an obligation to treat weak rational creatures decently, we can
(and undoubtedly would) call the knaves vile and odious. That we can condemn the knaves in
these terms shows that sentimentalism is in accord with ordinary moral thinking (Blackburn,
1998, p. 223). Nothing significant is lost. So contrary to the Kantian claim, we are not committed
to there being any categorical imperatives.

This objection seems misguided. Some of what is lost can be isolated by considering how a knave might respond to a sentimentalist condemnation of him: ‘You call my exploitation of the weak vile and odious. But I do not believe that it is, and I am under no obligation to change my behavior.’ On Blackburn’s account, all of what the knave says may well be true. But an advocate of categorical imperatives is able to hold that the last part of what he says is false. Contrary to what the knave contends, he is obligated to treat the weak decently. The sentimentalist gives up any hope of accusing the knave of uttering a falsehood, and thus clashes with ordinary moral thinking.

The sentimentalist seems to forfeit something else as well. The concept of obligation is closely linked to that of blame. A person would not morally blame another for failing to do something if he did not believe it to be morally obligatory to do it, it seems. For example, you would not blame me for taking a vacation if, in your view, I had no obligation to refrain from taking one. Of course, a person might be disappointed when another fails to do something, even if the person does not believe the other’s doing it to be obligatory. For example, someone who rejects the idea that donating a third of one’s income to charity is morally required might nevertheless regret that his very wealthy friend does not reach that level of giving: he might think she would be a better person if she did. But she would not be the object of his blame. A necessary condition for a person’s morally blaming another for doing something seems to be that he believe that refraining from doing it is morally obligatory. If this is correct, then, unlike an advocate of categorical imperatives, the sentimentalist cannot ever blame a foreign office knave for abusing outsiders. The sentimentalist’s evaluational capacities seem stunted. For in many
circumstances, we do hold foreign office knaves morally blamable. Granted, the sentimentalist can use some of the words we often employ in our judgments of foreign office knaves. He can call a knave odious and vicious. But he can call him this only in a sense that does not imply moral blame. We employ this sense when we apply the terms to bacteria and fungi. In the sentimentalist’s scheme the moral notion of blame disappears along with that of obligation in cases like that of the knave. And many of us believe this to be a genuine loss. It thus appears not only that the sentimentalist’s theory is incompatible with categorical imperatives, but also that its being incompatible with them has a significant cost.

A Case for Categorical Imperatives

Acknowledging these points would not doom the sentimentalist. For he can admit that his theory has some unfortunate implications, but maintain that it is the best we can do. It was understandable and even noble to try to show that morality incorporates categorical imperatives that are grounded ultimately in reason. But we can see now that it is a project doomed to failure (Blackburn, 1998, pp. 214-224).

I am unaware of any successful attempt to prove through philosophical argument that all of us are rationally compelled to embrace a particular principle as a categorical imperative. To my knowledge, no one has shown, for example, that self-contradiction is the price of denying Kant’s Formula of Humanity to be valid. In my view it is too early to give up on such attempts. But even if we do, we are not forced to abandon the position that morality is based in reason. We can hold that morality contains categorical imperatives, and that (along the lines of Kant’s second claim above) it is partly constitutive of being a rational agent to be bound by them.
Let me now try to illustrate the plausibility of holding this. A principle of reason, let us say, is any principle to which all rational agents ought to conform, unless doing so would violate some other, overriding principle to which all rational agents ought to conform. A categorical imperative, then, is a principle of reason that can never be overridden. Now to many of us, Kant seems on target in suggesting that the following is a principle of reason: ‘If you will an end, then will the means to it that are necessary and in your power’ (Kant, 1996b, p. 70). An agent would act contrary to this principle, sometimes called ‘the hypothetical imperative,’’ and thus irrationally by willing an end, say to lose weight, but failing to take the means he realizes to be necessary and in his power to attaining it, say significantly reducing his caloric intake. The dieter wills an end but, contrary to reason, fails to do what he understands he needs to in order to attain it. (Kant, I think, rightly rejects the idea that in all cases such as this the agent does not really will the end.) I know of no successful philosophical argument, that is, no successful argument that rests on premises that purport to be demonstratively irrational to deny, that proves irrationality to be the price of rejecting the hypothetical imperative as a principle of reason. It seems wrong-headed and even bizarre to deny (as Hume sometimes appears to) that there is ever any irrationality in willing an end but failing to will the means to it that one realizes to be necessary and in one’s power. Yet as far as I can tell it would be unwarranted at this point to embrace a conclusion as strong, say, as that anyone who denies this thereby falls into self-contradiction. Echoing Kant’s remarks regarding the moral law, it seems nevertheless that the hypothetical imperative is ‘firmly established of itself.’” A principle might be a principle of reason even though, strictly speaking, no philosophical argument demonstrates that it is.

Someone might acknowledge this point, and perhaps even agree that the hypothetical
imperative (or something like it) has a legitimate claim to be a principle of reason, but
evertheless express doubt as to whether there are any categorical imperatives. But consider
(what I call) the ‘principle of reasons universalism’: ‘If you judge an action permissible for a
rational agent in given circumstances, then you must (are rationally compelled) also to judge it
permissible for any other rational agent in relevantly similar circumstances.’ Despite being
hypothetical in grammatical form, this principle is a candidate for a categorical imperative. The
principle might make a non-overridable and unconditional claim on all of us. According to it, a
person must not judge that an action is permissible for him but not for another solely on the
grounds that he is he and the other is the other. If someone judges that an action is permissible
for himself but not for another, then it must be on the grounds that their circumstances, perhaps
including their personal qualities, somehow differ. It is very hard to envisage circumstances in
which the principle would be trumped by any other principle of reason. How, for example, could
it come into conflict with (and therefore be overridden by) an important, morally robust
categorical imperative like that not to treat persons merely as means? It seems plausible to hold
the principle of reasons universalism to be a categorical imperative, albeit a very modest one.
The principle would not prohibit much. As we have seen, it would not forbid a foreign office
knave from judging that it is permissible for him to exploit foreigners, as long as he is willing to
acknowledge that it is also permissible for foreigners to exploit him when they are in
circumstances analogous to his own.

Why focus on such a modest principle? It seems to have just as much claim to being a
principle of reason as the hypothetical imperative. But the principle of reasons universalism
appears to be a categorical imperative. So it is not as odd as one might think to believe in
It is even legitimate to wonder whether sentimentalism itself implicitly relies on a categorical imperative. According to Blackburn, ‘it is at least as natural to us pretty much to confine our cares to the family or tribe or other local group, as it is to expand our view to include sympathetic practical concern with everyone’ (Blackburn, 1998, p. 213). This seems to be correct. Yet recall that for Blackburn the common point of view has two components. It involves asking oneself whether, if an individual performed an action like this, it would have a positive impact on the individual’s family, tribe, or other local group. It also involves asking oneself whether one could see himself justifying this action to persons outside of his circle. If one answers either question negatively, then one will (supposedly) have a disagreeable sentiment towards the action and will presumably be obligated not to perform it. But why is this second component part of the common point of view? If, as Blackburn implies, for many if not most of us it is not natural to have sympathetic practical concern for everyone, why would we care whether we could see ourselves as justifying a course of action to outsiders? It would be unconvincing for the sentimentalist to reply simply that we have a pleasing sentiment towards the idea of our being able to justify ourselves to outsiders. For if we had this pleasing sentiment, then why would so many of us lack sympathetic practical concern for outsiders?

The sentimentalist might say that this aspect of the common point of view has its origins in self-interest (or group interest). Blackburn suggests that this is Hume’s response (Blackburn, 1998, p. 210). The idea seems to be that we cannot avoid interacting with outsiders, and to do well for our circle we need to see ourselves as able to justify our actions to them. This response is open to familiar objections. If we are devoid of sympathetic practical concern for a particular
group of outsiders, and they (or their agents) are unable to do us significant harm, then the promotion of our self-interest depends not at all on our ability to justify (or see ourselves as able to justify) our actions to them. Indeed, Hume suggests this point when he argues that we would not be obligated to be just to weak, rational creatures intermingled with us. Moreover, suppose that a different group of outsiders can do us significant harm and that we also have no sympathetic, practical concern for them. In order to promote our own interests, we might need to get them to believe that our actions towards them are justified. But we might accomplish this through deceit and manipulation, all the while caring not a whit that, in our own view, we are unable to justify our actions to them.

Objections such as these might be overcome, but I doubt it. So why might we care whether we are able to see ourselves as justifying our actions to outsiders? It is tempting to reply that it is because a principle of reason commands it. The principle might be a categorical imperative such as: ‘Treat rational agents only in ways that, in your view, you can justify to them.’ To me this seems every bit as plausible as claiming that what prompts us to care is self-interest alone. Of course, depending on the notion of justification employed, the implications of the imperative might be very weak. As we have noted, an agent might see himself as justified in treating an outsider in any way, as long as he acknowledges that, in similar circumstances, the outsider would be justified in treating him (the agent) in that way. On this notion of justification, fraud or coercion would not necessarily be ruled out.

My own view is that it is reasonable to believe in more robust categorical imperatives. A principle I take to be a plausible candidate is implicit in Kant’s Formula of Humanity. It is this: ‘Never treat rational agents merely as means.’ Very roughly, this imperative forbids any agent in
his interaction with another from behaving in a way to which the other cannot consent and, at the same time, pursuing an end which the other cannot share. Elsewhere I hope to offer a detailed defense of this principle. Doing so will, of course, involve making a case for the view that it possesses each feature of a categorical imperative. For example, it might seem that there are instances in which the principle would be overridden. Cases come to mind in which treating one person merely as a means seems to be morally permissible, since doing so is necessary to save the lives of many others. Part of the burden of defending the principle would be to show either that, despite initial appearances, these are not cases that violate it, or that, upon reflection, we are not committed in these cases to the view that violating the principle is morally permissible. If successful, this type of defense will show that it is reasonable to hold a particular principle to be a categorical imperative. It will not prove that inconsistency is the penalty for failing to do so.²

Samuel J. Kerstein

Samuel J. Kerstein is Associate Professor of Philosophy at the University of Maryland, College Park. He is working on a defense of the Kantian view that central to morality is the idea of humanity’s being an end in itself.
Notes

1. In another sense, however, an absolutely necessary principle might not be unconditional. That the imperative ‘If you are a parent, you must promote the welfare of your child’ contains a condition does not disqualify it from being absolutely necessary.

2. For a very different reading of *Groundwork* III, one that seems better grounded in Kant’s texts, see Schönecker (1999).

3. He is, of course, also invoking Hume’s ‘sensible knave,’ namely someone who ‘in particular incidents, may think that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any considerable breach in the social union and confederacy’ (Hume, 1975, 282).

4. Thanks to David Lefkowitz and Dieter Schönecker for help on this essay.
References


Further Reading


