A BRIEF INTRODUCTION TO KANTIAN ETHICS*

J. David Velleman
University of Michigan

The Overall Strategy

The overall strategy of Kant's moral theory is to derive the content of our obligations from the very concept of an obligation. Kant thought that we can figure out what we are obligated to do by analyzing the very idea of being obligated to do something. Where I am using the word 'obligation', Kant used the German word *Pflicht,* which is usually translated into English as "duty". In Kant's vocabulary, then, the strategy of his moral theory is to figure out *what our duties are* by analyzing *what duty is.*

A duty, to begin with, is a practical requirement—a requirement to do something or not to do something. But there are many practical requirements that aren't duties. If you want to read Kant in the original, you have to learn German: there's a practical requirement. Federal law requires you to make yourself available to serve on a jury: there's another practical requirement. But these two requirements have features that clearly distinguish them from moral obligations or duties.

The first requires you to learn German only if you want to read Kant in the original. This requirement is consequently escapable: you can gain exemption from it by giving up the relevant desire. Give up wanting to read Kant in the original and you can forget about this requirement, since it will no longer apply to you. The second requirement is also escapable, but it doesn't point to an escape hatch so clearly, since it doesn't contain an "if" clause stating a condition by which its application is limited. Nevertheless, its force as a requirement depends on the authority of a particular body—namely, the U.S. Government. Only if you are subject to the authority of the U.S. Government does this requirement apply to you. Hence you can escape the force of this requirement by escaping the authority of the Government: immunity to the authority of the body entails immunity to its requirements.

Now, Kant claimed—plausibly, I think—that our moral duties are inescapable in both of these senses. If we are morally obligated to do something, then we are obligated to do it no matter what our desires, interests, or aims may be. We cannot escape the force of the obligation by giving up some particular desires, interests, or aims. Nor can we escape the force of an obligation by escaping from the jurisdiction of some authority such as the Government. Kant expressed the inescapability of our duties by calling them **categorical** as opposed to **hypothetical**.

According to Kant, the force of moral requirements does not even depend on the authority of God. There is a simple argument for denying this dependence. If we were subject to moral requirements because they were imposed on us by God, the reason would have to be that we are subject to a requirement to do what God requires of us; and the force of this latter requirement, of obedience to God, could not itself depend on God's authority. (To require obedience to God on the grounds that God requires it would be viciously circular.) The requirement to obey God's requirements would therefore have to constitute a fundamental duty, on which all other duties depended; and so God's authority would not account for the force of our duties, after all. Since this argument will apply to any figure or body conceived as issuing requirements, we can conclude that the force of moral requirements must not depend on the authority of any figure or body by which they are conceived to have been issued. ~

The notion of authority is also relevant to requirements that are conditional on wants or desires. These requirements turn out to depend, not only on the presence of the relevant want or desire, but also on its authority.

Consider the hypothetical requirement "If you want to punch someone in the nose, you have to make a fist." One way in which you might escape the force of this requirement is by not wanting to punch anyone in the nose. But there is also another way. Even if you find yourself wanting to punch someone in the nose, you may regard that desire as nothing more than a passing fit of temper and hence as providing no reason for you to throw a punch. You will then regard your desire as lacking authority over you, in the

sense that it shouldn't influence your choice of what to do. The mere psychological fact that you want to punch someone in the nose doesn't give application to the requirement that if you want to punch someone in the nose, you have to make a fist. You *do* want to punch someone in the nose, but you *don't* have to make a fist, because the relevant desire has no authority.

All of the requirements that Kant called hypothetical thus depend for their force on some external source of authority—on a desire to which they refer, for example, or an agency by which they have been issued. And these requirements lack the inescapability of morality because the authority behind them is always open to question. We can always ask why we should obey a particular source of authority, whether it be a desire, the U.S. Government, or even God. But the requirements of morality, being categorical, leave no room for questions about why we ought to obey them. Kant therefore concluded that moral requirements must not depend for their force on any external source of authority.

Kant reasoned that if moral requirements don't derive their force from any external authority, then they must carry their authority with them, simply by virtue of what they require. That's why Kant thought that he could derive the content of our obligations from the very concept of an obligation. The concept of an obligation, he argued, is the concept of an intrinsically authoritative requirement—a requirement that, simply by virtue of what it requires, forestalls any question as to its authority. So if we want to know what we're morally required to do, we must find something such that a requirement to do *it* would not be open to question. We must find something such that a requirement would carry authority simply by virtue of requiring that thing. ~

Thus far I have followed Kant fairly closely, but now I am going to depart from his line of argument. When Kant derives what's morally required of us from the authority that must inhere in that requirement, his derivation depends on various technicalities that I would prefer to skip. I shall therefore take a shortcut to Kant's ultimate conclusion.

As we have seen, requirements that depend for their force on some external source of authority turn out to be escapable because the authority behind them can be questioned. We can ask, "Why should I act on this desire" or "Why should I obey the U.S. Government?" or even "Why should I obey God?" And as we observed in the case of the desire to punch someone in the nose, this question demands a reason for acting. The authority we are questioning would be vindicated, in each case, by the production of a sufficient reason.

What this observation suggests is that any purported source of practical authority depends on reasons for obeying it—and hence on the authority of reasons. Suppose, then, that we attempted to question the authority of reasons themselves, as we earlier questioned other authorities. Where we previously asked "Why should I act on my desire?" let us now as "Why should I act for reasons?" Shouldn't this question open up a route of escape from *all* requirements?

As soon as we ask why we should act for reasons, however, we can hear something odd in our question. To ask "Why should I?" is to demand a reason; and so to ask "Why should I act for reasons?" is to demand a reason for acting for reasons. This demand implicitly concedes the very authority that it purports to question—namely, the authority of reasons. Why would we demand a reason if we didn't envision acting for it? If we really didn't feel required to act for reasons, then a reason for doing so certainly wouldn't help. So there is something self-defeating about asking for a reason to act for reasons. ~

The foregoing argument doesn't show that the requirement to act for reasons is inescapable. All it shows is that this requirement cannot be escaped in a particular way: we cannot escape the requirement to act for reasons by insisting on reasons for obeying it. For all that, we still may not be required to act for reasons.

Yet the argument does more than close off one avenue of escape from the requirement to act for reasons. It shows that we are subject to this requirement if we are subject to any requirements at all. The requirement to act for reasons is the fundamental requirement, from which the authority of all other requirements is derived, since the

authority of other requirements just consists in there being reasons for us to obey them. There may be nothing that is required of us; but if anything is required of us, then acting for reasons is required.

Hence the foregoing argument, though possibly unable to foreclose escape from the requirement to act for reasons, does succeed in raising the stakes. It shows that we cannot escape the requirement to act for reasons without escaping the force of requirements altogether. Either we think of ourselves as under the requirement to act for reasons, or we think of ourselves as under no requirements at all. And we cannot stand outside both ways of thinking and ask for reasons to enter into one or the other, since to ask for reasons is already to think of ourselves as subject to requirements. ~

The requirement to act for reasons thus seems to come as close as any requirement can to having intrinsic authority, in the sense of being authoritative by virtue of what it requires. This requirement therefore comes as close as any requirement can to being inescapable. But remember that inescapability was supposed to be the hallmark of a moral obligation or duty: it was the essential element in our *concept* of a duty, from which we hoped that the *content* of our duty could be deduced. What we have now deduced is that the requirement that bears this mark of morality is the requirement to act for reasons; and so we seem to have arrived the conclusion that "Act for reasons" is the content of our duty. How can this be?

At this point, I can only sketch the roughest outline of an answer; I won't be able to supply any details until the end of this essay. Roughly, the answer is that to act for reasons is to act on the basis of considerations that would be valid for anyone in similar circumstances; whereas immoral behavior always involves acting on considerations whose validity for others we aren't willing to acknowledge. If we steal, for example, we take our own desire for someone else's property as a reason for making it our property instead—as if his desire for the thing wasn't a reason for its being his property instead of ours. We thus take our desire as grounds for awarding ownership to ourselves, while denying that his desire is grounds for awarding ownership for him. Similarly, if we lie,

we hope that others will believe what we say even though we don't believe it, as if what we say should count as a reason for them but not for us. Once again, we attempt to separate reasons for us from reasons for others. In doing so, we violate the very concept of a reason, which entails that a reason for one is a reason for all. Hence we violate the requirement "Act for reasons."

So much for a rough outline of Kant's answer. Before I can supply the details, I'll need to explore further what we feel ourselves required to do in being required to act for reasons. And in order to explore this requirement, I'll turn to an example that will seem far removed from morality.

Reasons that are Temporally Constant

Suppose that you stay in shape by swimming laps two mornings a week, when the pool is open to recreational swimmers. But suppose that when your alarm goes off this morning, you just don't feel like facing the sweaty locker room, the dank showers, the stink of chlorine, and the shock of diving into the chilly pool. You consider skipping your morning swim just this once.

(If you don't exercise regularly, you may have to substitute another example for mine. Maybe the exceptions that you consider making "just for this once" are exceptions to your diet, your drinking limit, or your schedule for finishing your schoolwork.)

When you are tempted to make an exception to your program of exercise, you are likely to search for an excuse—some reason for staying in bed rather than going off to the pool. You sniffle a few times, hoping for some signs of congestion; you lift your head to look out the window, hoping for a blizzard; you try to remember your calendar as showing some special commitment for later in the day. Excuse-making of this sort seems perfectly natural, but it ought to seem odd. Why do you need a reason for not doing something that you don't feel like doing?

This question can be understood in several different ways. It may ask why you don't already have a good enough reason for not swimming, consisting in the fact that you just don't feel like it. To this version of the question, the answer is clear. If not feeling like it were a good enough reason for not swimming, then you'd almost never manage to get yourself into the pool, since the mornings on which you're supposed to swim almost always find you not feeling like it. Given that you want to stay in shape by swimming, you can't accept "I don't feel like it" as a valid reason, since it would completely undermine your program of exercise. Similarly, you can't accept "That would taste good" as a reason for going over your limit of drinks, or you wouldn't really have a limit, after all.

Why not accept "I don't feel like it" as a reason on this occasion while resolving to reject it on all others? Again the answer is clear. If a consideration counts as a reason for acting, then it counts as a reason whenever it is true. And on almost any morning, it's true that you don't feel like swimming.

Yet if a reason is a consideration that counts as a reason whenever it's true, then why not dispense with reasons so defined? Why do you feel compelled to act for *that* sort of consideration? Since you don't feel like swimming, you might just roll over and go back to sleep, without bothering to find some fact about the present occasion from which you're willing to draw similar implications whenever it is true. How odd, to skip exercise in order to sleep and then to lose sleep anyway over finding a reason not to exercise! ~

Kant offered an explanation for this oddity. His explanation was that acting for reasons is essential to being a person, something to which you unavoidably aspire. In order to be a person, you must have an approach to the world that is sufficiently coherent and constant to qualify as a single, continuing point-of-view. And part of what gives you a single, continuing point-of-view is your acceptance of particular considerations as having the force of reasons whenever they are true.

We might be tempted to make this point by saying that you *are* a unified, persisting person and hence that you *do* approach practical questions from a point-of-view framed

by constant reasons. But this way of making the point wouldn't explain why you feel compelled to act for reasons; it would simply locate acting for reasons in a broader context, as part of what makes you a person. One of Kant's greatest insights, however, is that a unified, persisting person is something that you *are* because it is something that you *aspire to be*. Antecedently to this aspiration, you are merely aware that you are *capable* of being a person. But any creature aware that it is capable of being a person, in Kant's view, is *ipso facto* capable of appreciating the value of being a person and is therefore ineluctably drawn toward personhood.

The value of being a person in the present context is precisely that of attaining a perspective that transcends that of your current, momentary self. Right now, you would rather sleep than swim, but you also know that if you roll over and sleep, you will wake up wishing that you had swum instead. Your impulse to decide on the basis of reasons is, at bottom, an impulse to transcend these momentary points-of-view, by attaining a single, constant perspective that can subsume both of them. It's like the impulse to attain a higher vantage point that overlooks the restricted standpoints on the ground below. This higher vantage point is neither your current perspective of wanting to sleep, nor your later perspective of wishing you had swum, but a timeless perspective from which you can reflect on now-wanting-this and later-wishing-that, a perspective from which you can attach constant practical implications to these considerations and come to a stable, all-things-considered judgment.

If you want to imagine what it would be like never to attain continuing point-of-view, imagine being a cat. A cat feels like going out and meows to go out; feels like coming in and meows to come in; feels like going out again and meows to go out; and so on, all day long. The cat cannot think, "I have things to do outside and things to do inside, so how should I organize my day?" But when you, a person, find yourself to-ing and fro-ing in this manner, you feel an impulse to find a constant perspective on the question when you should "to" and when you should "fro".

This impulse is unavoidable as soon as the availability of the more encompassing vantage point appears. As soon as you glimpse the possibility of attaining a constant

perspective from which to reflect on and adjudicate among your shifting preferences, you are drawn toward that perspective, as you would be drawn toward the top of a hill that commanded a terrain through which you had been wandering. To attain that standpoint, in this case, would be to attain the single, continuing point-of-view that would constitute the identity of a person. To see the possibility of attaining it is therefore to see the possibility of being a person; and seeing that possibility unavoidably leads you to aspire toward it. ~

Of course, there is a sense of the word 'person' that applies to any creature capable of grasping the possibility of attaining the single, continuing perspective of a fully unified person. One must already be a person in the former, minimal sense in order to aspire toward personhood in the latter. I interpret Kant as having used words like 'person' in both senses, to denote what we already are and what we consequently aspire to become.

This Kantian thought is well expressed—believe it or not—by a word in Yiddish. In Yiddish, to call someone a *Mensch* is to say that he or she is a good person—solid, centered, true-blue. But *Mensch* is just the German word for "person" or "human being", like the English "man" in its gender-neutral usage. Thus, a *Mensch* in the German sense is merely a creature capable of being a *Mensch* in the Yiddish.

To be a solid, centered human being of the sort that Yiddishers call a *Mensch* entails occupying a unified, persisting point-of-view defined by a constant framework of reasons. But to be a human being at all, according to Kant, is to grasp and hence aspire toward the possibility of attaining personhood in this sense. Hence the imperative that compels you to look for generally valid reasons is an imperative that is naturally felt by all *Menschen:* the imperative "Be a *Mensch*".

The requirement "Be a *Mensch*" already sounds like a moral requirement, but I have introduced it by way of an example about exercise, which we don't usually regard as a moral obligation. My example may therefore seem ill suited to illustrate a requirement that's supposedly fundamental to morality. On second thought, however, we may have to reconsider what sort of a requirement we are dealing with.

If you do roll over and go back to sleep, in my example, you will be left with an emotion that we normally associate with morality—namely, guilt. You feel guilty when you shirk exercise, go over your drinking limit, put off working, or otherwise make an exception "just for this once." Indeed, your motives for seeking a reason on such occasions include the desire to avoid the sense of guilt, by avoiding the sense of having made a singular exception.

There is the possibility that the word 'guilt' is ambiguous, and that self-reproaches about shirking exercise do not manifest the same emotion as self-reproaches about lying or cheating. Alternatively, there is the possibility that the guilt you feel about shirking exercise is genuine but unwarranted. I would reject both of these hypotheses, however. If you go for your usual swim but stop a few laps short of your usual distance, you might well accuse yourself of cheating; if asked whom you were cheating, you would probably say that you were cheating yourself. Insofar as you owe it to yourself to swim the full distance, your sense of guilt may be not only genuine but perfectly appropriate.

Kant believed that moral obligations can be owed not only to others but also to oneself. Defenders of Kant's moral theory often seem embarrassed by his notion of having obligations to oneself, which is said to be odd or even incoherent. But I think that Kant's concept of an obligation is the concept of something that can be owed to oneself, and that any interpretation under which obligations to self seem odd must be a misinterpretation. That's why I have begun my account of Kantian ethics with self-regarding obligations.

Thus far, I have explained how the natural aspiration toward a stable point-of-view is both an aspiration to be a person, in the fullest sense, and a motive to act on considerations that have the same practical implications whenever they are true—that is, to act for reasons. I have thus explained how the felt requirement to be a person can deter you from cheating on your drinking limit or program of exercise and, in that minor respect, impel you to be a *Mensch*. What remains to be explained is how the same requirement can impel you to be a *Mensch* by eschewing other, interpersonal forms of cheating.

Reasons that are Universally Shared

In Kant's view, being a person consists in being a rational creature, both cognitively and practically. And Kant thought that our rationality gives us a glimpse of—and hence an aspiration toward—a perspective even more inclusive than that of our persisting individual selves. Rational creatures have access to a shared perspective, from which they not only see the same things but can also see the visibility of those things to all rational creatures.

Consider, for example, our capacity for arithmetic reasoning. Anyone who adds 2 and 2 sees, not just that the sum is 4, but also that anyone who added 2 and 2 would see that it's 4, and that such a person would see this, too, and so on. The facts of elementary arithmetic are thus common knowledge among all possible reasoners, in the sense that every reasoner knows them, and knows that every reasoner knows them, and knows that every reasoner knows that the every reasoner knows them, and so on.

As arithmetic reasoners, then, we have access to a perspective that is constant not only across time but also between persons. We can compute the sum of 2 and 2 *once and for all,* in the sense that we would only get the same answer on any other occasion; and each of us can compute the sum of 2 and 2 *one for all,* in the sense that the others would only get the same answer. What's more, the universality of our perspective on the sum of 2 and 2 is evident to each of us from within that very perspective. In computing the sum of 2 and 2, we are aware of computing it *for all,* from a perspective that's shared by all arithmetic reasoners. In this sense, our judgment of the sum is authoritative, because it speaks for the judgment of all.

This shared perspective is like a vantage point overlooking the individual perspectives of reasoners, a standpoint from which we not only see what everyone sees but also see everyone seeing it. And once we glimpse the availability of this vantage point, we cannot help but aspire to attain it. We are no longer satisfied with estimating or guessing the

sum of two numbers, given the possibility of computing it once for all: we are ineluctably drawn to the perspective of arithmetic reason.

Note that the aspect of arithmetic judgments to which we are drawn in this case resembles the authority that we initially regarded as definitive of moral requirements: it's the authority of being inescapable. We can compute the sum of 2 and 2 once for all because the answer we reach is the answer that would be reached from any perspective and is therefore inescapable. We can approach the sum of 2 and 2 from wherever we like, and we will always arrive at the same answer. The case of arithmetic reasoning shows that inescapability can in fact appeal to us, because it is the feature in virtue of which judgments constitute a stable and all-encompassing point-of-view. Perhaps, then, the authority of moral judgments, which consists in their inescapability, can appeal to us in similar fashion, by offering an attractive vantage point of some kind. ~

But what does arithmetic reasoning have to do with acting for reasons? Well, suppose that the validity of reasons for acting were also visible from a perspective shared by all reasoners—by all practical reasoners, that is. In that case, our aspirations toward personhood would draw us toward the perspective of practical reason as well.

Indeed, that may be the perspective toward which you were being drawn when you felt compelled to find a reason for not exercising. Your immediate concern was to find a set of considerations whose validity as reasons would remain constant through fluctuations in your preferences; but you would also have regarded those considerations as constituting reasons for other people as well, insofar as they were true of those people. In accepting an incipient cold as a reason to skip swimming, you would have regarded it as something that would count as a reason for anyone to skip swimming, in circumstances like yours. What you were seeking may thus have been considerations that could count as reasons not only for you, whenever they were true of you, but for other agents as well.

There is one important difference between practical and arithmetic reasoning, however. When you searched for reasons not to exercise this morning, no considerations just struck you as the ones that would strike any practical reasoner, in the way that 4 strikes you as

being the answer that would strike any reasoner adding 2 and 2. Rather, you had to try out different considerations as reasons; and you tried them out by testing whether you would be willing to have them strike you as reasons whenever they were true. That's how you tested and then rejected "I don't feel like it" as a reason for not exercising.

This feature of the case suggests that you may not have access to a pre-existing perspective shared by all reasoners in practical matters as you do in arithmetic. Apparently, however, you were trying to *construct* such a perspective, by asking whether you would be willing for various considerations to count as reasons whenever they were true, as if their reason-giving force, or validity, were accessible from a shared perspective. You asked, What if "I don't feel like it" were generally valid as a reason for not exercising?—as if you could choose whether or not to enshrine the validity of this consideration in a constant perspective of practical reasoning.

There is a sense in which you could indeed enshrine the validity of this consideration in a constant *individual* perspective. For if you had taken something as a sufficient reason for not exercising on this occasion, you would later have remembered doing so, and your deliberations on subsequent occasions might then have been guided by the precedent. Having once accepting a consideration as a reason for not exercising, you might later have felt obliged to accept it again, in other situations where it was true. Even so, however, you aren't capable of enshrining the validity of a consideration in a perspective that would be shared by all practical reasoners, since your taking something as a reason would not influence the deliberations of others as it would the deliberations of your future selves. Although you can construct a *temporally constant* perspective from which to conduct your own practical reasoning, you cannot construct a *universally shared* perspective. ~

And yet constructing a universally shared perspective of practical reasoning is precisely what Kant said that you must regard yourself as doing when you decide how to act. Kant expressed this requirement as follows: "Act only on a maxim that you can at the same time will to be universal law."

The clearest example of willing a maxim to be universal law—the clearest example that I know of, at least—is the train of thought that you undertake when considering whether to make an exception "just for this once," such as an exception to your diet or program of exercise. You think of potential reasons, in the form of true considerations such as "That would taste good" or "I don't feel like it," but then you realize that you aren't willing to grant these considerations validity as reasons whenever they are true, since doing so would completely undermine your regimen. Having found that you cannot consistently will these considerations to be generally valid as reasons, you refuse to act on them, as if in obedience to Kant's requirement.

According to Kant, however, you are required to act on considerations whose validity as reasons you can consistently will to be evident, not just to yourself on other occasions when they are true, but to other practical reasoners of whom they may be true as well. You are thus required to act only on considerations whose validity you could willingly enshrine in a universally accessible perspective of practical reasoning. That's what Kant meant by acting only on a maxim that you could will to be universal law.

Yet the force of Kant's proposed requirement remains elusive. Even if I have managed to direct your attention to your own sense of being required to construct a temporally constant perspective of practical reasoning, that requirement presupposes the possibility of your constructing such a perspective—a possibility that depends, in turn, on ties of memory between your current decision-making and your decision-making in the future. As we have seen, however, you aren't capable of constructing a perspective of practical reasoning that would be universally accessible to all reasoners. So how can you feel required to construct one?

I'm going to skip over the question for the moment, in order to describe how Kant's moral theory reaches its conclusions. I'll return to the question later, eventually offering two alternative answers to it. First, however, I want to show how substantive moral conclusions can issue from Kant's theory.

Two Examples

Suppose that we were required to act only on considerations whose validity as reasons we could willingly enshrine in a universally accessible perspective of practical reasoning, just as we feel required to act only on considerations whose validity we could enshrine in a temporally constant perspective. This requirement would decisively rule out some considerations. Here is an example from Kant's *Critique of Practical Reason:*

> Suppose, for example, that I have made it my maxim to increase my fortune by every safe means. Now, I have a deposit in my hands, the owner of which is dead and has left no writing about it. This is just the case for my maxim. I desire then to know whether that maxim can also hold good as a universal practical law. I apply it, therefore, to the present case, and ask whether it could take the form of a law, and consequently whether I can by my maxim at the same time give such a law as this, that everyone may deny a deposit of which no one can produce a proof. I at once become aware that such a principle, viewed as a law, would annihilate itself, because the result would be that there would be no deposits. [*Critique of Practical Reason* 27]

In this passage, Kant imagines considering whether a consideration such as "I want the money" can count as a reason for denying the receipt of a deposit from someone who has died without leaving any record of it. Much as you asked whether you were willing to make "I don't feel like it" valid as a reason for not exercising on all occasions when it is true, Kant asks whether he is willing to make "I want the money" valid as a reason for all trustees of whom it is true. Kant says, "The result would be that there would be no deposits." Why not?

The answer is that the validity of reasons for denying unrecorded deposits would have to be common knowledge among all practical reasoners. If a trustee's desire to keep a depositor's money were a valid reason for denying its receipt, then the validity of that reason would have to be known to prospective depositors, who have access to the common knowledge of practical reasoners, and who would then be deterred from making any deposits, in the first place. A trustee can therefore see that he would never receive a single deposit if wanting to keep it would be a valid reason for him to deny its receipt, just as the drinker sees that he wouldn't have a limit if his thirst were a valid reason for exceeding it.

A trustee can therefore see that if "I want the money" were a valid reason for denying the receipt of deposits, there would be no deposits whose receipt he could deny. And a consideration can hardly be a reason for an action that would be rendered unavailable by the validity of that very reason. "I want the money" couldn't be a universally accessible reason for defaulting because, if it were, there would be no opportunities for defaulting. And since it couldn't be a universally accessible reason, it isn't be valid as a reason for defaulting, after all. ~

Actually, this example is an instance of a larger class, since defaulting on the return of a deposit would unavoidably involve lying, and lying also violates the fundamental requirement "Act for reasons." So let's examine this larger class of examples.

To lie is intentionally to tell someone a falsehood. When we tell something to someone, we act with a particular kind of communicative intention: we say or write it to him with the intention of giving him grounds for believing it. Indeed, we intend to give him grounds for belief precisely by manifesting this very communicative intention in our speech or writing. We intend that the person acquire grounds for believing what we say by recognizing that we are acting with the intention of conveying those grounds.

Now, suppose that our wanting to give someone grounds for believing something constituted sufficient reason for telling it to him, whether or not we believed it ourselves. In that case, the validity of this reason would be common knowledge among all reasoners, including him. He would therefore be able to see that, in wanting to give him grounds for believing the thing, as was manifest in our communicative action, we already had sufficient reason for telling it to him, whether or not we believed it. And if he could see that we had sufficient reason for telling it even if we ourselves didn't believe it, then our telling it would give him no grounds for believing it, either. Why should he believe what we tell him if we need no more reason for telling him than the desire, already manifest in the telling, to give him grounds for believing it? So if our wanting to give him grounds for believing something were sufficient reason for telling it to him, then telling him wouldn't accomplish the result that we wanted, and wanting that result

wouldn't be a reason for telling him, after all. Wanting to convey grounds for belief can't be a sufficient reason for telling, then, because if it were, it would not be a reason at all. ~

I introduced these example by asking you to imagine that you could construct a universally accessible perspective of practical reasoning, so that you could be required to act only on considerations whose validity you could enshrine in such a perspective. Yet it has now turned out that there already *is* such a perspective—or, at least, the beginnings of one—and it hasn't be constructed by anyone. For we have stumbled on one kind of practical result that anyone can see, and can see that anyone can see, and so on.

The kind of practical result that we have found to be universally accessible has the following form: that the validity of some putative reason for acting could not be universally accessible. The validity of "I want the money" as a reason for denying receipt of a deposit, or the validity of "I want him to believe it" as a reason for telling something to someone, could not be universally accessible, any more than the validity of "That would taste good" as a reason for going over your limit of drinks. The fact that the validity of these reasons could not be universally accessible—*this* fact is already universally accessible to practical reasoners, any of whom can perform the reasoning by which it has come to light.

Thus, the notion of sharing a perspective with all practical reasoners is not a pipe-dream, after all. You already share a perspective with all practical reasoners to this extent, *that it is common knowledge among all reasoners that the validity of certain reasons for acting could not be common knowledge among all reasoners*. This item of common knowledge constitutes a universally accessible constraint on what can count as a reason for acting and hence what can satisfy a requirement to act for reasons. A requirement to act for reasons would forbid acting on the basis of considerations whose validity as reasons could not be common knowledge among all reasoners, and in the case of some considerations, this impossibility is itself common knowledge. ~

Let me review the argument to this point, which can now be seen to implement the

overall strategy of deriving the content of our duties from the very concept of a duty. We began with the idea that moral requirements must be inescapable, which led to the idea that they must be intrinsically authoritative, in the sense of having authority over us simply by virtue of what they require. We then found a requirement that came as close as possible to having such authority—the requirement to act for reasons, which cannot coherently be questioned and must be presupposed by all other practical requirements.

Next we saw how the requirement to act for reasons is experienced in ordinary life, when one looks for an exemption from some regular regimen or policy. In this example, the requirement to act for reasons is experienced as an impulse to act on a consideration from which one is willing to draw the same consequences whenever it is true, an impulse that militates against cheating oneself. And we found such an impulse intelligible as part of one's aspiration toward the unified, persisting point-of-view that makes for a fully integrated person.

Our next step was to observe that rational creatures can attain not only unified individual perspectives but a single perspective that is shared, in the sense that its deliverances are common knowledge among them. And with the help of examples drawn from Kant, we saw that a requirement to act on considerations whose validity was common knowledge would amount to a ban on cheating others. What remains to be explained is how the requirement to act for reasons in this sense is experienced in ordinary life and whether it, too, can be understood as part of the aspiration to be a person.

The Idea of Freedom

In order to answer this remaining question, we must return to a problem that we considered earlier and set aside—the question why we feel compelled to think of ourselves as constructing a universally accessible framework of reasons for acting. We can't actually build a universally accessible framework of reasons, although we do enjoy universal access to the fact that some reasons, in particular, couldn't be built into such a

framework. The question is why we feel compelled not to act on reasons that couldn't be built into something that isn't for us to build, in the first place.

Kant's answer to this question was that in order to act, we must conceive of ourselves as free; and that in order to conceive of ourselves as free, we must conceive of ourselves as acting on reasons that owe their authority to us. Considerations have authority as reasons only if they have the sort of validity that is universally accessible to all reasoners; but we won't be free in acting on them, Kant believed, if they have simply been dictated to us from a universal perspective in which we have no say. We must think of them as reasons on which we ourselves confer authority, by introducing them into that perspective. ~

I think that Kant was simply wrong about the idea of freedom, insofar as he thinks that it requires us to be the source of the authority in our own reasons for acting. Roughly speaking, I think that we cannot be guided by reasons whose only authority is that with which we ourselves have endowed them.

To endow reasons with authority, as I have now conceived it, would be to *make* their status as reasons common knowledge among all reasoners—a feat that is simply beyond our power. More importantly, it's a feat that we cannot help but *think* is beyond our power. If we thought that something's being a reason could become common knowledge among all reasoners only by dint of our making it so, then we would have no hope of its ever being so. Hence if we thought that reasons owed their authority to us, we would have no hope of their ever having authority.

Why can't reasons owe their authority to us? The answer is that endowing reasons with authority would entail making their validity common knowledge among all reasoners. And if we could promote reasons to the status of being common knowledge among all reasoners, then we should equally be able to demote them from that status—in which case, the status wouldn't amount to rational authority. The point of a reason's being common knowledge among all reasoners, remember, is that there is then no way of evading it, no matter how we shift our point-of-view. No amount of rethinking will make such a reason irrelevant, because its validity as a reason is evident from every

perspective. But if we could decide what is to be common knowledge among all thinkers, then a reason's being common knowledge would not entail its being inescapable, since we could also decide that it wasn't to be common knowledge, after all. Our power to construct a universally accessible framework of reasons would therefore undermine the whole point of having one.

I think that Kant's mistake was to claim that we must act under the idea of **freedom**; what he should have said, I think, is that we must act under the idea of **autonomy**. Let me explain the difference between these concepts.

'Autonomy' is derived from the Greek word for self-rule or self-governance. Our behavior is autonomous when it is self-governed, in the sense that we ourselves are in control of it; it is not autonomous—or, as Kant would say, it is **heteronomous**—when it is controlled by something other than ourselves. To say that behavior is controlled by something other than ourselves is not to say that it is controlled from outside our bodies or our minds. A sneeze or a hiccup is not under our control; neither is a startle or an impulsive cry of pain; but all of these heteronomous behaviors originate within us. What makes them heteronomous is that, while originating *within,* they don't originate *with us:* they aren't fully our doing. Only the behaviors that are fully our doing qualify as autonomous actions.

The fact that we act autonomously doesn't necessarily entail that we have free will—not, at least, in the sense that Kant had in mind. In Kant's view, our having free will would require not only that we sit behind the wheel of our behavior, so to speak, but also that we face more than one direction in which it would be causally possible that we steer it, so that our future course is not pre-determined. One might suspect that if our future course *was* pre-determined, then we wouldn't really be in control of our behavior, and hence that autonomy really does require freedom. Yet there is a way for us to follow a pre-determined course and yet steer that course in a meaningful sense. Our course might be pre-determined by the fact that there are reasons for us to do particular things and that we are rationally responsive to reasons. So long as we are responding to reasons, we remain autonomous, whether or not those reasons pre-determine what we do.

Consider here our autonomy with respect to our beliefs. When we consider the sum of 2 and 2, we ourselves draw the conclusion that it is 4. The thought 2+2=4 is not dictated to us by anyone else; it is not due to an involuntary mental association, not forced on our minds by an obsession or fixed in our minds by a mental block; in short, it isn't the intellectual equivalent of a sneeze or a hiccup. When we consider the sum of 2 and 2, we make our own way to the answer 4. And yet there is no other answer that we could arrive at, given that we are arithmetically competent and that, as any reasoner can see, the sum of 2 and 2 is 4. So when we consider the sum of 2 and 2, we are pre-determined to arrive at the answer 4, but to arrive there autonomously, under our own intellectual steam. We aren't free to conclude that 2+2 is 5, and yet we are autonomous in concluding that it is 4.

Perhaps, then, we can steer our behavior as we steer our thoughts, in directions that are pre-determined, not by exogenous forces, but by our rational ability to do what there is reason for doing, just as we think what there is reason for thinking. In that case, we could have autonomy without necessarily having free will. ~

Kant himself identified what is special about behavior that is rationally necessitated. Whereas heteronomous behavior is determined by antecedent events under a law of nature, he observed, autonomous behavior is determined by *our conception* of a law. A law, in this context, is just a practical requirement of the sort with which this analysis of duty began, a requirement specifying something that we must do. What makes our behavior autonomous is that we do it, not just because our doing it is necessitated by prior events, but because we realize that doing it is required—a realization that constitutes our conception of a law, in Kant's terms. Our recognition of a practical requirement, and our responsiveness to that recognition, is what makes the resulting action attributable to us, as our doing: it's what gets us into the act.

Kant thus explained why acting for reasons makes us autonomous. Acting for reasons makes us autonomous because "Act for reasons" is the ultimate requirement lying behind all other practical requirements, whose authority depends on there being reasons to obey them. Whenever our behavior is determined by our conception of law—that is, by our

realization that some action is required—we are being governed at bottom by a recognition of reasons, either constituting or backing up that requirement.

Kant thought that being determined by our recognition of a practical requirement, on the one hand, and being determined by prior events under a law of nature, on the other, are mutually exclusive alternatives, at least in the sense that we cannot conceive of ourselves as being determined in both ways at once. (In fact, he thought that we can perhaps *be* determined in both ways at once but that we can't *conceive* of being so, because we can't reconcile these two modes of determination in our minds.) But I think that being determined by our recognition of a practical requirement can itself be conceived as a causal process, governed by natural laws. I express this possibility by saying that we can conceive of ourselves as autonomous without having to conceive of ourselves as free. ~

Because Kant thought that we cannot conceive of ourselves as autonomous without also conceiving of ourselves as free, he insisted that we must not conceive of practical requirements as externally dictated. That is, we must not find ourselves confronted with inexorable reasons for doing things, in the way that we find ourselves confronted with an inexorable answer to the calculation of 2+2; for if we did, our action would be predetermined, and we wouldn't be free to choose it, just as we aren't free to choose a sum for 2+2. Kant thought that we must regard the balance of reasons for acting as being up to us in a way that the sum of 2 and 2 is not.

Kant's insistence that we act under the idea of freedom thus led him to insist that we conceive of ourselves as constructing rather than merely finding a universally accessible framework of reasons for acting. As I have explained, I think that our constructing reasons would deprive them of the authority that universal accessibility is meant to provide. But as I have also explained, I think that Kant's insistence on our constructing them is unnecessary, because we can act under the idea of autonomy, without any pretensions of being free.

Even if we need only think of ourselves as autonomous when we act, we will still be required to act for reasons, since autonomy consists in being determined by authoritative considerations. The requirement to act for reasons can thus be felt to arise from the aspiration to be a person in a more profound form. Our earlier discussion directed our attention toward the general region of experience where the requirement to act for reasons can be found, but it didn't identify the fundamental manifestation of that requirement. We saw that the requirement to act for reasons can be felt to arise from our aspiration to be a person, but we traced it to a fairly specific instance of that aspiration, consisting in our aspiration toward a temporally constant point-of-view. And then we found that this specific aspiration cannot account for moral force of the requirement in interpersonal cases. The present discussion suggests that the fundamental manifestation of the requirement to act for reasons is a different form of the aspiration to be a person: it's the aspiration toward autonomy. We feel required to act for reasons insofar as we aspire to be persons by being the originators of our own behavior ~

Contradictions in the Will

Replacing Kant's references to freedom with references to autonomy needn't alter our analysis of the foregoing examples. The aspiration toward autonomy yields a requirement to act for reasons, and this requirement will forbid us to act on considerations whose practical implications couldn't be common knowledge, as in the cases of cheating analyzed above.

Yet there are other cases in which Kant derived moral conclusions in a way that depends on the very aspect of freedom by which it differs from what I have called autonomy. In these examples, what rules out some considerations as reasons for acting, according to Kant, is not they couldn't be universally accessible, as in the case of our grounds for stealing or lying, but rather that we couldn't consistently *make* them universally accessible. It is precisely our inability to build these considerations into a universally accessible framework of reasons that prevents them from being reasons, according to Kant. Yet our inability to build some considerations into a universally

accessible framework of reasons would prevent them from being reasons only if such a framework depended on us for its construction—which is what I have just been denying, in contesting Kant's view of freedom. My disagreement with Kant on the subject of freedom therefore threatens to escalate into a disagreement about which considerations can be reasons and, from there, into a disagreement about what is morally required.

The clearest cases of this kind have the form of prisoner's dilemmas. Prisoner's dilemmas get their name from a philosophical fiction in which two people—say, you and I—are arrested on suspicion of having committed a crime together. The police separate us for interrogation and offer us similar plea bargains: if either gives evidence against the other, his sentence (whatever it otherwise would have been) will be shortened by one year, and the other's sentence will be lengthened by two. The expected benefits give each of us reason to testify against the other. The unfortunate result is that each sees his sentence shortened by one year in payment for his own testimony, but lengthened by two because of the other's testimony; and so we both spend one more year in jail than we would have if both had kept silent.

Let me pause to apologize for a misleading feature of this story. Because the characters in the story are criminals, and the choice confronting them is whether to tell the whole truth and nothing but the truth, turning state's evidence may seem to be the option that's favored by morality. But this story serves as a model for every case in which the choice is whether to join some beneficial scheme of cooperation, such as rendering aid or keeping commitments to one another. There are parts of morality whose basic point is to enjoin cooperation in cases of this kind, and philosophers use the prisoners' dilemma as a model for those parts of morality. In order to understand philosophical uses of the prisoners' dilemma, then, we have to remember that cooperating with one's fellow prisoner represents the moral course in this philosophical fiction, because it is the course of mutual aid and commitment.

Prisoners' dilemmas are ripe for Kantian moral reasoning because the two participants are in exactly similar situations, which provide them with exactly similar reasons. When each of us sees the prospect of a reduced sentence as a reason to testify against the other,

he must also see that the corresponding prospect is visible to the other as a reason for doing likewise, and indeed that the validity of these reasons is common knowledge between us.

Given that our reasons must be common knowledge, however, I ought to wish that the incentives offered to me were insufficient reason for testifying against you, since the incentives offered to you would then be insufficient reason for testifying against me, and both of us would remain silent, to our mutual advantage. And you must also wish that the incentives were insufficient reason for testifying against me, so that I would likewise find them insufficient for testifying against you. Furthermore, each of us must realize that the other shares the wish that the incentives were insufficient reason for turning against the other. The following is therefore common knowledge between us: we agree in wishing that what was common knowledge between us was that our reasons for turning against one another were insufficient. ~

Here the power to construct a shared framework of reasons would certainly come in handy, since you and I would naturally converge on which reasons to incorporate into that framework and which reasons to exclude. The power to construct a shared framework of reasons would thus transform our predicament, in way that it would not have transformed the cases considered earlier.

In the case of lying, for example, we found that it was not just undesirable but downright impossible that our desire for someone to believe something should be a sufficient reason for telling it to him. This desire couldn't possibly be such a reason, we concluded, because its being a reason would entail common knowledge of its being one, which in turn would ensure that it wasn't a reason, after all. This conclusion did not depend on the assumption that we could in any way affect the rational import of wanting someone to believe something—that we could elevate it to the status of a reason or demote it from that status. Even if reasons were handed down to us from a universally accessible perspective that we took no part in constructing, we would know in advance

that the deliverances of that perspective would not include, as a sufficient reason for telling something to someone, the mere desire that he believe it.

Hence our conclusion about lying is not at all threatened by the doubts outlined above about the Kantian doctrine of freedom. But those doubts do threaten the prospect of drawing any Kantian conclusions about the prisoners' dilemma. For whereas some reasons for lying are rendered impossible by the necessity of their being common knowledge, our reasons for turning against one another in the prisoners' dilemma are rendered merely undesirable. And if reasons are indeed handed down to us from a universally accessible perspective that we take no part in constructing, then we have no guarantee against being handed undesirable reasons, even if they were universally undesirable. Only if we construct the shared framework of reasons can we expect it to exclude undesirable reasons, such as our reasons for turning against one another in the prisoner's dilemma. ~

Our proposed reasons for lying are ruled out by what Kant called a **contradiction in conception**. This contradiction prevents us from conceiving that the desire for someone to believe something should be a sufficient reason for telling it to him. Kant thought that our proposed reasons for turning against one another in the prisoners' dilemma can also be ruled out, not because a contradiction would be involved in their *conception,* but rather because a contradiction would be involved in their *construction*—a contradiction of the sort that Kant called a **contradiction in the will**. Specifically, building these reasons into the universally accessible framework would contradict our desire that what was common knowledge between us were reasons for cooperating instead. But if the framework of reasons is not for us to construct, then contradictions in the will are no obstacle to anything's being a reason, and half of Kantian ethics is in danger of failure. Securing Kantian ethics against this failure requires a substantial revision in the theory, in my opinion. I'll briefly outline one possible revision.

The prisoners' dilemma places you and me at odds not only with one another but also with ourselves. If you find that the incentives are a sufficient reason for turning state's

evidence, you will wish that they weren't, given that their status as a reason must be common knowledge between us, which will persuade me to turn state's evidence as well. You therefore find yourself in possession of reasons that you wish you didn't have. Of course, you may often find yourself in such a position. As you drag yourself out of bed and head for the pool, for example, you may wish that you didn't have such good reasons for sticking to your regimen of exercise. These cases may not involve any contradiction in your will, strictly speaking, but they do involve a conflict, which complicates your decision-making and compromises the intelligibility of your decisions. Think of the way that you vacillate when confronted with unwelcome reasons for acting, and the way that you subsequently doubt your decision, whatever it is.

I have argued that you cannot simply will away unwelcome reasons for acting, but the fact remains that you can gradually bring about changes in yourself and your circumstances that mitigate or even eliminate the conflict. You can learn to relish early-morning swims, you can switch to a more enticing form of exercise, or you can find some other way to lower your cholesterol. You can also cultivate a disdain for advantages that you wouldn't wish to be generally available, such as the advantages to be gained in the prisoner's dilemma by turning against a confederate. You might even learn to regard an additional year in prison as badge of honor, when it is incurred for refusing to turn against a confederate, and a shortened sentence as a mark of shame under these circumstances—in which case, the plea bargain offered to you would no longer be a bargain from your point-of-view, and the prisoners' dilemma would no longer be a dilemma. This attitude towards incarceration can't be called up at a moment's notice, of course; it may take years to cultivate. But when you adopted a life of crime, you could have foreseen being placed in precisely the position represented by the prisoners' dilemma, and you could already have begun to develop attitudes that would clarify such a position for you. (Surely, that's what lifetime criminals do, and rationally so—however irrational they may be to choose a life of crime, in the first place.)

Thus, if you find yourself confronted with unwelcome reasons for acting, you have probably failed at some earlier time to arrange your circumstances or your attitudes so as to head off conflicts of this kind. You can't change your personality or your

circumstances on the spot; nor can you change their status as reasons for acting here and now. But with a bit of foresight and self-command, you could have avoided the predicament of acting on reasons that you wished you didn't have. Since you had reason for taking steps to avoid such a conflict, you have somewhere failed to act for reasons—not here and now, as you act on your unwelcome reasons, but at some earlier time, when you allowed yourself to get into that predicament.

Hence the requirement "Act for reasons" can favor morality in two distinct ways. First, it can rule out various actions, such as lying, that are based on considerations whose validity as reasons is inconceivable. Second, it can rule out acquiring reasons whose validity, though conceivable, is unwelcome. In the latter case, it doesn't rule out performing any particular actions; rather, it rules out becoming a particular kind of person, whose reasons for acting are regrettable, even from his own point-of-view.

Before I turn from the current line of thought, I should reiterate that it cannot be traced to the works of Kant himself. Kant would reject the suggestion that contradictions in the will are always such as to have occurred long before the time of action, and hence to be beyond correction on the spot. The resulting moral theory is therefore kantian with a small *k*.

Respect for Persons

There is one more way in which the requirement to act for reasons constrains us to be moral, in Kant's view. Kant actually thought that this constraint is equivalent to the ones that I've already discussed—that it is one of the aforementioned contradictions viewed from a different angle or described in different terms. I disagree with Kant on this point, and so I'll present this constraint as independent of the others, thus departing again from Kant.

Many people take up a regimen of diet or exercise as a means of staying healthy, but some overdo it, so that they ruin their health instead. Most people accumulate money as a means of buying useful or enjoyable things, but some overdo it, grubbing for money so

hard that they have no time to spend it. In either case, the over-doers are making a fundamental mistake about reasons for acting: they are exchanging an end for the means to that end, thus exchanging something valuable for something else that is valuable only for its sake. Exercise is not valuable in itself but only for the sake of health (or so I am assuming for the moment); money is not valuable in itself but only for the sake of happiness. To sacrifice health for the sake of exercise, or to sacrifice happiness for the sake of money, is to stand these values on their heads. The prospect of gains in exercise or income can't provide reason for accepting a net loss in the ends for whose sake alone they are valuable.

Kant's greatest insight, in my view, was that we can commit the same mistake in practical reasoning with respect to persons and their interests. The basis of this insight is that the relation between a person and his interests is similar to, though not exactly the same as, the relation between an end, such as happiness, and the means to it, such as money. Kant believed that persons themselves are ends, and that they consequently must not be exchanged for the things that stand to them in the capacity analogous to that of means. ~

Some commentators interpret Kant as meaning that persons are ends in the same sense as health or happiness—that is, in the sense that we have reason to promote or preserve their existence. What Kant really meant, however, is that persons are things *for the sake of which* other things can have value.

The phrase 'for the sake of' indicates the subordination of one concern to another. To want money for the sake of happiness is to want money because, and insofar as, you want to be happy; to pursue exercise for the sake of health is to pursue it because, and insofar as, you want to be healthy. You may also care about things for the sake of a person. You may want professional success for your own sake, but you may also want it for the sake of your parents, who love you and made sacrifices to give you a good start. In the latter case, your concern for your happiness depends upon your concern for others; in the former, it depends upon your concern for yourself.

The dependence between these concerns is evident in the familiar connection between how you feel about yourself and how you feel about your happiness. Sometimes when you realize that you have done something mean-spirited, you come to feel worthless as a person. You may even hate yourself; and one symptom of self-hatred is a loss of concern for your own happiness. It no longer seems to matter whether life is good to you, because you yourself seem to be no good. Your happiness matters only insofar as *you* matter, because it is primarily for your sake that your happiness matters at all.

Now, to want money for the sake of happiness is to want the one as a means of promoting or preserving the existence of the other; but to want happiness for your own sake is not to want it as a means of promoting or preserving your existence. Happiness is not a means of self-preservation, and the instinct of self-preservation is not the attitude that underlies your concern for it. The underlying self-concern is a sense of your value as a person, a sense of self-worth, which is not at all the same as the urge to survive. Hence wanting happiness for your own sake is both like and unlike wanting money for the sake of happiness. The cases are alike in that they involve the subordination of one concern to another; but they are unlike with respect to whether the objects of concern are related as instruments and outcomes. ~

When Kant referred to persons as ends, he was not saying that they lend value to anything that stands to them as instruments, or means. He was saying merely that they are things for the sake of which other things can have value, as your happiness is valuable for your sake. The dependence between these values, however, is enough to yield a rational constraint similar to the constraint on exchanging ends for means.

If your happiness is valuable for your sake, and matters only insofar as you matter, then you cannot have reason to sacrifice yourself for the sake of happiness, just as you cannot have reason to sacrifice happiness for the sake of money. Just as your concern for money is subordinate to your concern for happiness, so your concern for happiness is subordinate to self-concern, and the former concerns must not take precedence over the

latter, as would happen if you pursued money at the sacrifice of your happiness, or happiness at the sacrifice of yourself.

Sacrificing yourself for the sake of happiness may sound impossible, but it isn't. People make this exchange whenever they kill themselves in order to end their unhappiness, or ask to be killed for that purpose. The requirement to act for reasons rules out such mercy killing, which exchanges a person for something that's valuable only for his sake. Because a person's happiness is valuable for his sake, it cannot provide a reason for sacrificing the person himself.

(Before I go further, I should point out that Kantian ethics does not, in my view, rule out suicide or euthanasia in every case. As we have seen, Kantian ethics rules out actions only insofar as they are performed for particular reasons. For example, it doesn't rule out false utterances in general but only those which are made for the sake of getting someone to believe a falsehood. Similarly, it doesn't rule out suicide and euthanasia in general but only when they are performed for the sake of ending unhappiness. With that qualification in place, let me return to my explanation of persons as ends.) ~

Kant thought that the status of persons as ends rules out more than sacrificing them for their interests; he thought that it rules out treating them in any way that would amount to using them merely as means to other ends. In his view, persons shed value on other things, by making them valuable for a person's sake; whereas means merely reflect the value shed on them by the ends for whose sake they are valuable. To treat a person as a means is to treat him as a mere reflector of value rather than a value-source, which is a confusion on the order of mistaking the sun for the moon. Indeed, Kant thought that a universe without persons would be pitch dark with respect to value.

Here let me remind you of the aspiration in which the requirement to act for reasons is manifested in our experience. Reasons for acting are considerations that are authoritative in the sense that their practical import is common knowledge among all reasoners, including not only other people but also ourselves at other times. Having access to such considerations enables us to act autonomously, as the originators of our own behavior.

And being autonomous is essential to—perhaps definitive of—being a person. Hence the requirement to act for reasons expresses our aspiration to realize a central aspect of personhood—or, as I put it, the aspiration to "be a *Mensch*."

This alternative formulation of the requirement to act for reasons has implications for the current discussion of persons as ends-in-themselves. What it implies is that the felt authority of reasons is due, in part, to our appreciation of ourselves as persons. In acting for reasons, we live up to our status as persons, and we act for reasons partly as a way of living up to that status. The motivational grip that reasons have on us is subordinate to our appreciation for the value of being a *Mensch*.

If you think back to our initial search for an intrinsically inescapable requirement, you will recall that "Act for reasons," though close to being inescapable, was not perfectly so. We settled for it after reflecting that we are required to act for reasons if we are subject to any requirements at all. What we have subsequently discovered is that seeing ourselves as subject to practical requirements is essential to seeing ourselves as autonomous and, in that respect, as persons. Thus, although we are required to act for reasons only insofar as we are subject to practical requirements at all, we are obliged to conceive of ourselves as subject to requirements, and hence required to act for reasons, by our aspiration toward personhood. ~

The value of persons now emerges as paramount, not only over the value of what we do for someone's sake, but over the value of acting for any reason whatsoever. Acting for reasons matters because being a person matters.

What's more, the value of our individual personhood here and now is inseparable from the value of participating in personhood as a status shared with our selves at other times and with other people, whose access to the same framework of reasons is what lends those reasons authority. Only by sharing in the common knowledge of reasoners do we find ourselves subject to authoritative requirements, recognition of which must determine our behavior if we are to be autonomous persons. Being an autonomous person is thus impossible without belonging to the community of those with access to the same sources

of autonomy. Insofar as being a person matters, belonging to the community of persons must matter, and the importance of both is what makes it important to act for reasons.

That's why it's irrational to treat any person merely as a means, for any reason whatsoever. No reason for acting can justify treating a person as a mere reflector of value, because the importance of acting for reasons depends on the importance of personhood in general as a source of value. Reasons matter because persons matter, and so we cannot show our regard for reasons by showing disregard for persons.

---

*This essay is an attempt to reconstruct Kantian moral theory in terms intelligible to undergraduates who have not yet read Kant. In the interest of commending to students those parts of Kant's theory which seem right to me, I have changed parts that seem wrong, usually with an explanation of my reasons for doing so. I have also chosen not to complicate the essay with references either to the Kantian texts or to the secondary literature, although my debts to others are numerous and not always obvious. I am especially indebted to the work of Elizabeth Anderson, Michael Bratman, Stephen Darwall, Edward Hinchman, Christine Korsgaard, and Nishi Shah.