



*Christine M. Korsgaard: Internalism
and the Sources of Normativity*

This is the version of the interview with Professor Korsgaard that was supposed to have appeared in *Constructions of Practical Reason: Interviews on Moral and Political Philosophy*, edited by Herlinde Pauer-Studer (Stanford: Stanford University Press, 2002). Due to an unfortunate accident, the first edition of that volume contains an unedited transcript of that interview rather than the corrected version below.

HERLINDE PAUER-STUDER: You have been a pupil of John Rawls. What influence did Rawls's thinking have on you? What do you think Rawls's most lasting contribution to practical philosophy has been, and what relations and differences are there between your own work and Rawls's work?

CHRISTINE M. KORSGAARD: First of all, a lot of Rawls's influence on me and on most of his students has been methodological. What I would like to say is, to put it a little bluntly, that he taught us how to read a book. Most of Rawls's students were deeply impressed with the approach that he took to the classics of philosophy, which was to insist that we read them not to criticize them, but to learn from them. He urged us to interpret them generously and to make them say something as reasonable and interesting as possible. I feel very strongly that he taught us how to learn from the history of philosophy, and that he put me and many other people on the path of learning from the history of philosophy. He did that with respect to Kant, of course, but he taught us how to learn from other philosophers as well. Of course, another influence is that he helped to create my great respect for Kant. In particular, I have been influenced by the

broadly Kantian account of how people should relate to one another which characterizes Rawls's political philosophy as well as Kant's moral philosophy.

In regard to his contributions to practical philosophy: I think that Rawls's political philosophy will be counted among the great works of political philosophy and is of enduring value. In particular, in his later work he has produced the first really consistent and thorough account of what liberalism involves. So I think he has made an essential contribution to political philosophy. His contribution to ethics more generally I have already specified by mentioning the influence he has had on me and so many other people. Rawls has influenced the way ethical philosophy is now done by teaching us all how to regard the tradition of ethics and how to learn from it.

As far as relations and differences between my work and Rawls's work: I suppose the main difference is that I have been inclined to focus on questions of individual morality and personal relationships rather than on questions of political philosophy. This is partly for the very natural reason that when you are someone's student and your teacher's achievement in political philosophy has been so massive, you are not inclined to do political philosophy yourself. You tend to think, "That has already been done." You want to take the same sort of insights and methods into another area. So there is a difference of emphasis, but it is not a matter of disagreement.

H. P.-s.: In your paper "Rawls and Kant: On the Primacy of the Practical"¹ you argue that practical philosophy with its specific way of problem solving could be a paradigm for philosophy altogether. Besides Kant you mention John Rawls as a thinker using this practical way of reasoning—namely that a certain position is justified because it is the only answer available to a certain problem, as the comparison between the content of the problem and the content of the answer makes clear. Could you elaborate in more detail on this primary role of practical philosophy and how your own work fits into this picture?

C. M. K.: I see philosophy as being primarily addressed to normative questions, which arise not only in moral and political philosophy but in every domain of philosophy. And I think that the problems of philosophy always or at least very often take the form of trying to understand why some purported normative claim really is normative. You are faced with a philosophical problem because of some circumstance you are in, in which some norm seems to apply to you, and you want to see whether it does or not. And I think that the most powerful way of establishing a normative claim often takes this form: You show your audience that because of some circumstance they are in, they have a certain normative problem, that is, they need a principle; and then you show them

that something is the best or the only solution to that problem, and then they are committed to that solution. And the solution is the normative principle.

Now, the problem itself can be some universal human problem or plight. In my work the problem is posed by the fact that we are self-conscious, and therefore need reasons for belief and action. And I mean to be reasoning from the fact that we are self-conscious and need reasons for action to our commitment to the moral law. In Rawls's work the problem is posed by something less universal but just as real, which is the problem of designing a liberal society, the problem of finding principles that people who disagree about nearly everything else can share. But in both cases the reasoning is supposed to go from a problem, which your reader is supposed to acknowledge that he or she has, to the best or only solution to that problem. And it is the fact that it is the best or only solution to the problem that gives the claim normative force.

That is a bit schematic, but I think that many philosophical problems have this form. In Kant's theoretical philosophy, for example, the problem is posed to us by the fact that we are confronted with a mass of phenomena that we have to reconstruct as a unified world in which we can find our way around. So it can be a deep kind of problem like that. I would oppose this way of thinking about philosophy to a model of philosophy that is very common and misleading, and that sees philosophy as being a lot like science, only more speculative. This scientific model sees the aim of philosophy as explanatory and supposes that when we do philosophy we are seeking some *theory* we could then apply; for instance, a theory about which actions are right. I think that is the wrong model for pursuing normative questions.

H.P.-s.: In current moral philosophy there is often drawn a sharp contrast between an Aristotelian type of moral theory and a Kantian type of moral theory. Your work on Aristotle and Kant suggests that—despite some differences—there are strong similarities between the two philosophers. Would you say that a synthesis between Kant and Aristotle is the way to a richer conception of morality answering the criticisms raised against Kant's ethics by defenders of so-called virtue ethics?

C.M.K.: I would say that. I have never been able to see why people think that Aristotle's philosophy and Kant's philosophy are opposed. I myself see very deep affinities between them. Both of them think that ethics is an expression of practical rationality. Both of them think about the mind in terms of the distinction between its active and its passive elements or aspects. The emphasis on that distinction—more specifically, the view that reason is the active element and sensibility is the passive element and that they have to work together in get-

ting a grip on the world—is a powerful common element in the philosophy of mind of these two thinkers. I think it is also very much at work in their ethics. Both of them see the human mind as being in an important way active, as making it possible to transcend mere reactivity in our relation to the world. And in a certain way they both think that this is what ethics is about.

Both of them also acknowledge a distinction I have been thinking about a lot lately, which is the distinction between (as I put it) an action and a mere act. The idea is that an action—what I mean is something that is captured by Kant's idea of a maxim—is an act for the sake of a certain end. To use a Kantian example, 'telling a lie' would be an act; 'telling a lie in order to get some money from a would-be lender' would be an action. According to both Aristotle and Kant it is the whole thing—this act for the sake of that end—that is the object of choice. And both of them think that actions in that sense are the units of moral assessment and the bearers of moral value.² This sets them apart from many other moral philosophers. In the case of consequentialism, it is obvious that the unit of moral value is the act—and this is why it seems so obvious to consequentialists that it is productive force, effects, which constitute moral value. But even in eighteenth-century forms of deontology, where the question was about intrinsic rightness, the emphasis was often on the acts rather than on actions. So I see the focus on actions, in this sense, and the view that they are the bearers of moral value as a deep affinity between Aristotle and Kant.

I think of Aristotle as giving us a theory of the virtues. The term "virtue" can be used as a general synonym for being a good person. But it can also be used in a more specific sense, which I think is at work in both Aristotle and Kant, to refer to the features of sensibility that a human being must have in order to be receptive to the demands of practical reason. Virtue in this sense is the perfection of the sensible, that is, the passive or receptive, side of our nature. I think that Aristotle gives us a brilliant theory of virtue in that sense and a very subtle moral psychology to go with it. Kant, on the other hand, gives us a better account of the basis and content of the principles of practical reason. But I think that each of these accounts is almost completely compatible with the other and together they amount to a very powerful theory.

H.P.-s.: According to your interpretation Kant's conception of morality is clearly a virtue ethic invoking moral reflection from a first-person perspective. Such a view—though I would agree that it is the correct interpretation of Kant—inevitably raises the well-known objection that a Kantian ethic reduces to a mere *Gesinnungsethik*, that it is confined to a single-person point of view and that it misses the intersubjective dimension of morality. Now, you have ar-

gued—using Wittgenstein’s private language argument—that there is nothing like “private reasons,” that all reasons coming up in the course of moral reflection are public or shared reasons.

One might still object that the mere linguistic expression of x and the argumentation about x are not the same, that there is a difference between public language and moral discourse. Would not the latter amount to an invitation of others and their point of view into the realm of moral reasoning and moral discussion and would this not result in giving up a first-person perspective?

C.M.K.: First of all, about the arguments I make which are based on Wittgenstein’s private language argument: I take the private language argument to provide not so much a route to the publicity of reasons as a sort of analogy for the publicity of reasons. So it is not merely the publicity of language itself that is supposed to do the moral work here. The analogy goes something like this: The words of a language have to have a public standing in order to have in some identifiable way the same meaning from one time to the next as we use them. And in a similar way I think reasons have to have the same force from one time to the next as we appeal to them. So reasons have a kind of public normative force of their own. It is not just a matter of the publicity of the language in which we speak about reasons. Their normativity itself is public.

I think that one way to understand the idea that reasons have this kind of public normative force is in terms of the theory of personal identity, on which I hope at some point to do more work. Behind the idea that reasons are private, the idea that they belong to one person in particular, is the assumption that we can identify one person in particular in advance of or prior to the person’s reasons themselves. Before I can say which reasons are “mine” there must be a “me.” But I hold a constructive view of personal identity, and I think that what gives us a personal or practical identity is the reasons that we autonomously adopt for ourselves. And that means that you cannot take for granted the notion of somebody’s being “me” in advance of which reasons I have and adopt. My reasons are part of my practical identity, and I construct my practical identity and so I do not exist, so to speak, prior to my reasons. And since I do not exist prior to my reasons, there is no way to identify a set of private reasons. Reasons have to be public because my reasons are part of what holds me together over time. And this is similar to the way in which the meanings of words have to be public for the language to be the same over time. This might sound a bit abstract, but that is the basic idea behind reasons being public.

H.P.-S.: Would then not the others have to come in and would this not break up the first-person perspective?

С.М.К.: I don't think so. What you get as a result of this is essentially the public and conversational nature of reasons. I do not think it brings in a third-person perspective. I think that what it does is shift us to a plural first-person perspective, which is very different from a third-person perspective. I associate the third-person perspective with dealing with people as objects, as parts of the phenomenal world. It is the point of view from which we are concerned with scientific explanation. The plural first-person perspective is a perspective, in which we are concerned with justification, rather than explanation. It is a perspective, in which we are joined with others in the processes of deliberation and justification; so in it we deal with people not as objects but as fellow subjects. In this form Wittgensteinian ideas are very harmonious with Kantian ones.

Н.Р.-С.: If you identify morality with a virtue ethics, the question arises how you cover the ground that is not captured by a virtue ethics, namely questions of public morality like justice and issues of political morality. Would you follow here Kant's partition between *The Doctrine of Virtue* (Tugendlehre) and *The Doctrine of Right* (Rechtslehre), and would you think this gives all the theoretical notions we need to deal with issues of morality in the personal as well as the public realm?

С.М.К.: I would not quite say that I regard Kant's conception of morality as a virtue ethic. It is true that I think that the primary notion in Kantian ethics is a notion of who you should be or what you should be like. The central value is a condition of the agent, namely autonomy—not in the sense that it is the most important thing, but in the sense that it is the value from which all other values arise. But I do not much like the distinction of ethical theories into virtue ethics, consequentialism, and deontology, because I think that any complete moral theory must include an account of all of the central moral notions, and must organize them in some systematic way. So I would not say that Kant's is a virtue ethics.

I do agree with Kant's partition between *The Doctrine of Virtue* and *The Doctrine of Right*. To explain why I'd like to go back to the ideas we were discussing at first about seeing philosophy as solving problems. You could see *The Doctrine of Virtue*, or ethics, and *The Doctrine of Right*, or politics, as being addressed to two different problems. The problem of ethics is how we are to act given that we have free wills and therefore must choose our own principles of action; the problem of politics is how we can be free in a world in which we interact with others. Kant sees these two problems as arising from two different domains of freedom: inner freedom of the will and outer freedom or liberty of action. The need to find principles that express your inner freedom is the prob-

lem addressed in the doctrine of virtue; the need to coordinate everybody's outer freedom in a way that maintains that freedom is the problem addressed in the doctrine of right. These two problems exist side by side, and have related but different solutions. Although the two domains need to be systematically related, neither of them has to be dependent on or be a branch of the other. I think there is one way in which *The Doctrine of Right* does not cover everything we want to say about political life. We want to say something not only about the laws that formally govern our relations but about the kind of community that a political unit forms. But I think there is room for this in a Kantian account.

I see Kant as deeply indebted to Rousseau in his political philosophy and also in his account of personal relationships. And I see this indebtedness to Rousseau as related to what I said a moment ago about the possible role of the idea of the plural subject in Kant's philosophy. A state is a kind of plural subject: the idea of the general will, which Kant borrows from Rousseau, is the idea of a shared will among a number of people. In Kant's account of personal relations we also find an emphasis on the idea of forming a shared will with someone, of having a bond of love or friendship. One way to look at it is this—morality involves the will we share with anyone just in virtue of our common human nature; politics involves the will we share with those with whom we live together on a shared territory; and personal relations involve the wills we share with those to whom we have particular connections. All of those things exist side by side and are separate domains of normative problems, and solutions, and resulting obligations.

H.P.-s.: Some philosophers argue that a moral theory has to have at least a consequentialist structure (without being necessarily utilitarian), including a means-end conception of rationality in order to be able to take the consequences of actions into account. It has been a well-known objection (if we think, for example, of Max Weber's distinction between *Gesinnungsethik* and *Verantwortungsethik*) that Kant's ethic does not take consequences into account. What do you think of this objection and how does your modified Kantian moral theory answer it?

C.M.K.: Certainly I do not think that a moral theory has to have a consequentialist structure. Earlier I mentioned, as a common point between Kant and Aristotle, the view that the unit of moral assessment is the action, the act undertaken for the sake of a certain end, rather than merely the act by itself. Acts may be assessed primarily in terms of their consequential value, but actions, the units of moral value, should not be. Of course I do not think that it is correct to say that Kantian agents do not care about or are not interested ei-

ther in the consequences of their acts. It would be impossible even to formulate a maxim without attention to the intended consequences of an act. So, I think, there is in a way a very deep disagreement here about what the unit of assessment is.

Consequentialism, as I see it, is not actually a moral theory because it supposes that the value of “doings” just amounts to the productive value of acts—that all we have to care about is the effects of our acts. And that is refusing to assign value to actions as such. So, if you think with Kant and Aristotle that what morality is all about is actions as a whole, then consequentialism doesn’t seem to be a moral theory. Consequentialism is a kind of technological vision, something proposed as a replacement for morality. It is a social engineering project.

But actions are also events in the world (or correspond to events in the world, at least), and they too have consequences. There are a number of different ways in which one can deal with worries about what happens to the consequences in Kant’s ethical theory. It is worth pointing out that Kant himself not only did not ignore the consequences, but took the fact that good actions can have bad effects as the starting point for his religious philosophy. In his religious thought, Kant was concerned with the question how the moral agent has to envision the world, how he has to think of its metaphysics in order to cope with the fact that the actions morality demands may have terrible effects that we never intended, or may simply fail to have good ones.

I myself see the development of what Rawls has called “nonideal theory” to be the right way of taking care of a certain class of cases, in which the consequences of doing the right thing just seem too appalling for us to simply wash our hands of. But I do not want to say that just having bad consequences is enough to put an action into the realm of nonideal theory. I think there is a range of bad consequences that a decent person has to be prepared to live with, out of respect for other people’s right to manage their own lives and actions, and to contribute to shared decisions. But I also think that there are cases where our actions go wrong in such a way that they turn out in a sense not to *be* the actions we intended to do, or to instantiate the values we meant them to instantiate. I think that some of these cases can be dealt with by introducing the kind of double-level structure into moral philosophy that I have described in the essay on “The Right to Lie: Kant on Dealing with Evil.”³ But I also think there are cases that cannot be domesticated even in this way, cases in which, to put it paradoxically, the good person will do something “wrong.” I have written about that sort of case too, in “Taking the Law into Our Own Hands: Kant on the Right to Revolution.”⁴

H.P.-S.: Your answer to the problem of the source of normativity is that our will, our capacity to reflect the normative status of our actions is the source of moral obligation. The crucial question is about how we understand ourselves; at stake is our practical identity, that is, whether we want to be moral subjects acting only on maxims on which all rational beings in a cooperative system could agree. You point out that this notion of practical identity amounts to a full conception of who we are, including desires, inclinations, and passions.

Do you think that this way the notorious problem of a dualism between reason on the one side and inclination, desires, and passions on the other dissolves? The question is also whether the introduction of this concept of practical identity does not relativize your conception in regard to Kant's. Will not—depending on their varying contingent identities—people give themselves different laws, and will not their varying powers of reflection lead to different results?

Could you elaborate more on how the role of rational reflection or reflective endorsement works in the account of moral obligation? Are our reflective capacities generally a reliable guide to make us aware of our obligations and of necessitation?

C.M.K.: As far as the dualism between reason on the one side and inclinations, desires, and passions on the other side is concerned, I do not think that the dualism completely dissolves, and I do not think we want it to dissolve. I mentioned earlier that both Kant and Aristotle see a certain set of philosophical problems as raised by the fact that the mind is both passive and active; they ask how the mind is to work in the face of that fact. I think that in any reasonable theory of moral psychology there has to be room for something like a kind of passive input that is an original spur to action, what Kant calls an incentive. But the agent is active in choosing to act on this incentive, and the form of the agent's activity is expressed in his principle. So an action is motivated by an incentive that is actively endorsed and chosen in light of some principle. It is an essential part of this picture that there should be some element of passive input.

I think some of the attacks on Kant's view of inclinations, desires, and passions are based on a kind of misunderstanding. To say that there are original inputs that are natural and in face of which we are passive is not necessarily to say that the kinds of things we ordinarily call desires, which may have a lot of cognitive content in them, are the same as these original inputs. If you think of these original inputs as being themselves what we ordinarily mean by "desires," then you get this picture, which people actually attribute to Kant: A desire is a thing in the face of which you are passive that somehow washes over you, a force of nature that comes from outside of you and is essentially an unintelligent

thing. All desires are more or less like animal appetites. And then reason either has to fight the desire or choose to allow it to operate.

There is a much more complicated view, which Kant himself starts to sketch in “Conjectural Beginnings of Human History,” in which reason works on the passive inputs and develops more intelligent, cognitively-loaded, specifically human desires from them. So there is room for a certain complexity in the intermingling of passive and active elements in the formation of the more developed states that we ordinarily call “desires.” It is essential to my picture, though, both that there be some passive natural inputs into the development of these desires and that those inputs explain the fact that the desires operate as incentives. This is partly just because, I think, that is how it works. But it is also because, I think, it is important to Kant’s theory of value to maintain the connection between what we are, in virtue of our nature, *capable* of valuing and the more formed values that actually come out of the system. As I read Kant, his view is that human beings create values, but we don’t create them from nothing and therefore we can’t value just anything. We create them from the resources of our natural psychology and that is what provides the limitations. So, as I said before, I do not want to get rid of the dualism altogether.

I don’t intend that the notion of practical identity should make my conception of obligation more relativistic than Kant’s. In *The Sources of Normativity* I make an argument that everyone who reflects must ultimately come to see her humanity itself as an essential and foundational feature of her practical identity. And it is this form of practical identity that is supposed to give us moral obligations in the strict sense, things we owe to other human beings simply as human beings. This is a conception of ourselves that we should be able to reach, so to speak, reasoning backwards from any particular conception of practical identity, regardless of which one it is, and asking why it is normative for us. There is a sense in which the conception of ourselves as valuable-qua-human is like an Idea of Reason in Kant’s philosophy. The unconditional Ideas of Reason are ideas that you work back to through a process of inference regardless of what particular phenomenon you start from. And then they are universal in their import and normative force. And I see the conception of the value of one’s humanity as being like that. So that is supposed to be what makes it a universal and nonrelativistic conception. Obviously I also have to argue that moral obligation follows from seeing oneself as valuable-qua-human in order to make good on that claim. But that is the basic idea of the argument.

As to the question whether our reflective capacities are generally a reliable guide, and make us aware of our obligations: There are, I think, two ways in which our reflective capacities fall short of being a completely reliable guide.

One of them is that we might not reflect far back enough. Kantian positions in general set a high value on reflection and are idealizing positions in the sense that moral concepts, as Kant defines them, are derived from the ideal of a fully reflective person. The fully reflective person is a corollary of Kant's idea of the unconditioned. We seek the unconditioned by imagining a person who reasons all the way back, who never gives up until there is a completely undeniable, satisfying, unconditional answer to the question. Obviously human beings often stop reflecting very far short of that. And reflection itself is not the solution to that problem. So in that sense reflection is not the complete guide although it is the only place where we can find guidance. Something else has to get us to reflect.

The other way, or maybe a related way, in which reflection is not a complete guide has to do with some of the things I said earlier about Aristotle. Aristotle, I said, gives us a theory of virtue in the sense that he gives us a theory about what our passive or sensible nature has to be like in order for our active nature to function properly. One of the most important ramifications of the fact that we have both passive and active faculties is this: No matter what problems thinking solves for you, something has to get you to think and that is not thinking itself. Thinking always starts from a stimulus, something from outside, just as choice always starts from a stimulus, something from outside.

And that is why in his later writings Kant gives us the beginnings of a theory of the virtues that is more like Aristotle's. In *The Doctrine of Virtue* Kant is concerned with the way in which our ends shape our sensible faculties so that we will recognize the occasions in which virtuous action is called for.

H.P.-s.: You basically follow Kant in the justification of the Categorical Imperative, although you make the details of the argument much clearer. The crucial step is that a free will has to give itself a law (otherwise it would not be free), and the only thing that is required of this law is that it has to be a law, which feature (that is, the lawlike form) is exactly fulfilled by the Categorical Imperative (in its first formulation). For Kant the justification of the Categorical Imperative in its different formulas is thus completed as Kant assumes the identity of the different formulas. But you yourself give up the claim of the identity of the different formulas. You think, for example, that the Formula of Humanity and the Formula of Universal Law lead to different results. Thus the problem arises that the justification for the Categorical Imperative in the Universal-Law formulation does not automatically justify the Categorical Imperative in the Formula-of-Humanity version. How would you close the gap in the justification? Could you elaborate a bit more on the work the different formulations of the Categorical Imperative do in dealing with concrete moral problems?

C.M.K.: I think that we need to distinguish among three different things. First, there is the Formula of Universal Law in the purely *formal* sense, which in *The Sources of Normativity* I called the Categorical Imperative. Then, there is the Formula of Universal Law in the sense of a principle, which demands that we act on reasons that we can share with all rational beings who live together in a cooperative community, which in *The Sources of Normativity* I called the Moral Law. And then, there is the Formula of Humanity. In *The Sources of Normativity* I argued that the foundational argument in fact only gets us to the formal version of the Formula of Universal Law, the Categorical Imperative. Before we can know what that principle requires of us, we need to specify what the principle universalizes over. The argument I described in answer to your question about relativism—the argument that is supposed to show us that we have to see our human identity as a normative form of practical identity—is supposed to answer that question; it shows us that what the Formula of Universal Law universalizes over is human beings as ends in themselves. So by that route what the foundational argument actually brings us to most immediately is something more like the Formula of Humanity.

The difference I pointed out between the Formulas of Universal Law and Humanity in the paper “The Right to Lie” is more a difference between what I just now called the Moral Law—the Formula of Universal Law seen as universalizing over rational agents who live cooperatively—and the Formula of Humanity. The Formula of Humanity turns out to be stronger than the Formula of Universal Law even in that more substantive sense. So it is not so much that there is a gap in the justification as simply that the Formula of Universal Law in the more substantive sense is still a more limited, and in a way a more social, principle than the Formula of Humanity. The Formula of Humanity gives us not only our duties to others but also our duties to ourselves. As far as the justification of morality goes, as I said a minute ago, I think the best argument goes directly from the formal principle to the Formula of Humanity. But I have to admit that this is a set of questions about which I am constantly changing my mind.

As far as dealing with concrete moral problems goes, I am very much inclined to the view, and have been for a long time, that the Formula of Humanity is the best formula to work with in dealing with concrete moral problems. I think that the formal version of the Formula of Universal Law, the Categorical Imperative, is essential to Kant’s overall theory because of the role it plays in the justificatory argument. But as far as the more substantive version of the Formula of Universal Law is concerned: I think it has a role as an aid to moral thinking

but I might give it a somewhat more reduced role than Kant had in mind for it. Certainly I do not think you can always arrive at the right answer to a moral question by asking whether you can will your maxim as a universal law without contradiction. But I do think that asking whether you can will your maxim without contradiction is a way of uncovering morally relevant and important features of our maxims.

For instance, there is a well-known set of problems about coordination, in which it turns out that you cannot will a maxim as a universal law not because the action is wrong, but because some coordination problem would arise if everyone acted in the same way. You know the sort of thing I mean: not everyone could adopt the maxim "I will become a professor just because I want to" if *in fact* everyone wanted to. So is it wrong to become a professor just because you want to? Some people use that sort of thing as an argument against the Formula of Universal Law. I am rather inclined to think that it shows us an interesting way of thinking about the role of the Formula of Universal Law in thinking about moral issues. If you run into a coordination problem the Formula of Universal Law makes you aware that that is the kind of human situation you are in, the kind of situation in which people have to coordinate their activities, and also that that is a *morally* important fact. So I suppose that means that I think of the Formula of Universal Law as a heuristic guide in moral thinking. But the Formula of Humanity is more powerful; it is more likely actually to give us correct moral answers.

H.P.-S.: The often neglected issue of moral motivation has lately received more attention by moral philosophers. You offer along Kantian lines an answer to the problem of moral motivation: A good-willed person acts on certain reasons, and the reason why the action is done is the same as the reason why the action is right. You nevertheless add that this account of moral motivation only works if we assume that people are rational and open to the force of rational considerations.

The objection might now be raised that this strong presupposition of rationality would also close the gap in an externalist account between a principle that is right and our being motivated by it, as we just need the assumption that cognitive considerations have the power to motivate.

C.M.K.: I would not quite say that the account of moral motivation only works if we assume that people are rational. Rather, I would say that there is a descriptive sense in which people have no choice but to be rational and to act on reasons of some kind. Rationality in this descriptive sense is forced upon us by the fact that we are self-conscious beings and can act on our incentives only

if we take them to be reasons. So there is no question of acting rationally versus not acting rationally. There is only a question whether our reasons are good ones or bad ones, whether we are rational in a normative sense. (And of course there is the precedent question whether we can derive some standard for reasons being good or bad ones, such as the Formula of Humanity in the argument I described before.) So I don't think that I am making a strong presupposition of rationality in the normative sense. It is more a thesis in moral psychology.

I think that the externalist position you are imagining is simply that being rational in this sense, being self-conscious, somehow forces us to be receptive to cognitive considerations ("reasons") whose normative force comes from completely outside of us. The form of the receptivity would be that when we know them we respond appropriately. The trouble with that position is that it is too blank, it doesn't explain anything, why we respond that way or what makes the response appropriate. This is related to criticisms that I have made of realist positions in general. You can of course just say, "Since we are rational beings, if there are any reasons out there we have to be responsive to them." But this does not give us a way of explaining what the real connection between the human rational being and the nature of reasons is. Let me put it this way: there is a way in which what someone who uses that strategy does is to take the notion of "a reason" to be the basic notion; then he simply defines rationality as receptivity to those items. There are reasons or rational principles out there, and when you respond to them in the right way, then you are rational. What I want to do is take the notion of our rational nature as being the basic notion and define reasons in terms of it. The first thing to ask is what it is about us that makes reasons exist in the world. I think it is our self-consciousness. So the connection between reasons and rationality goes the opposite way in the picture that I am talking about than it does in the externalist picture.

H.P.-S.: Your answer to Bernard Williams's position—that in order that we can be motivated by x , x has to be part of our motivational set and that pure reason cannot be part of our motivational set—is that rational considerations can be part of the motivational set. One might object that by this move you have assumed but not really proved that rational considerations motivate us. Does not this conception exclude again that things like inclinations and desires can motivate us morally?

How would you answer these criticisms and could you elaborate in more detail on your conception of moral motivation?

C.M.K.: Given what I just said about the order of explanation between rationality on the one hand and reasons or rational principles on the other, I prob-

ably would not now respond to Williams's argument in exactly the same way that I did when I wrote "Skepticism about Practical Reason."⁵ I still hold the position I described there, but the way I wrote that paper makes it possible to confuse my position with the one I just rejected—that rational principles are somehow just out there and you are rational if you respond to them correctly. I think that there is something misleading about the way Williams sets up the question. He sets up the subjective motivational set as something with some items in it, over here, so to speak, and the reasons that emerge from deliberation over there, and then he says: there has to be some path between them. The implicit assumption is that the principles of practical reason are transmitters of motivational or normative force from one reason to another. I don't think that is what principles of practical reason are. I think they are principles that determine what *counts* as a reason. To put it in the terms I used a moment ago: as I see Kant's moral psychology, every reason involves two elements—an incentive and a principle under which that incentive is chosen. If we did not have some principle of treating our desires as reasons, then desires would not be part of the subjective motivational set in a rational being. In Kant's philosophy the relevant principle is the principle of self-love, which is, as he sees it, a kind of basic animal tendency to take your natural inclinations to be reasons to act. So there must be some principle that makes desires and inclinations part of the subjective motivational set, part of the will, insofar as you can translate Kant's ideas into these terms at all. If we think of the principles of practical reason in this way then the focus of our attention is going to be on those principles and which ones they are, and, to get back to the point I was just making, we are going to find that out by thinking about what rationality is, what it means to be a rational being. That is the interesting question because it determines what rational principles there are, and so what is in the subjective motivational set and what is not. If rational principles determine what is in the subjective motivational set, then of course there will be a connection between the items in the set and the outputs of rational deliberation, and so internalism will be true, but that is a trivial result.

H.P.-S.: Starting from Kant's Formula of Humanity you develop a value theory, namely that it is *we* as rational beings, as ends in themselves, that confer value on things. This amounts to a constructive account assuming that values are only brought into the world via human beings. The clear advantage of this position is that it avoids dubious ontological assumptions about "objectively given" values, but it raises the question whether it does justice to what is at stake in issues of value.

In questions of value the unanimity of judgment is especially hard to reach, and the reference to features of the object seems an indispensable objective basis. How would you counter the objection that your account of values is too subjectivist, missing the relevant features in the objects that make something valuable? (For example, when we say a picture is valuable we would not say that it is valuable because we constructed its value, but because it has certain objective features which make it valuable.)

C.M.K.: Well, first of all, features of the objects certainly play a role in the value theory as I understand it because, as I mentioned earlier, we are not capable of valuing just anything. We can value something only if it has some form of natural appeal to us, where “natural appeal” is to be understood broadly—it satisfies one of our needs, it is the kind of thing we are capable of being interested in, it may have aesthetic appeal. Something in us has to answer to something in the object in order for the object to be valuable. Admittedly, though, I do not think that is much of a limitation because I think in Kant’s view part of what is involved in valuing humanity is valuing humanity’s—let me say it in an extreme way—creative power with respect to the making of values. It is an important feature of human beings, one that I think Kant wishes us to celebrate, that human beings are capable of taking an interest in, and therefore making values out of, just about anything. In some of my papers I have emphasized some of our more trivial values because they express this power: for example, people who collect ordinary objects like coins or stamps. The great variety of different traditions there are in human arts in different cultures also expresses the fecundity of the human power to create values.

Insofar as there is objectivity to values, insofar as we have to share them, I think this is not so much a fact about the valuable objects as it is a fact about us and our relations to each other. Roughly speaking, the idea is that understanding your own creative relation to your values should give you a basis for understanding the creative relation that other people stand in to their values. Trying to see what other people find of value in whatever they care about is a way of being rightly related to those people. So the sharing of values, as far as I am concerned, comes not from something internal to the valued objects, but rather from the fact that part of what it is for human beings to stand in a proper relation to one another is that we appreciate one another’s capacities, minds, tastes, interests, creativity.

H.P.-S.: In connection with your value theory you attack the usual opposition between “instrumentally good things” and “intrinsically good things.” You argue instead for a fourfold distinction, namely between, on the one hand,

“intrinsically good” (things which have their value in themselves) and “extrinsically good” (things which derive their value from some other source) and, on the other hand, between “ends or final goods” (things valued for their own sake) and “instrumental goods or means” (things valued for the sake of something else). This leads to the assumption that there is only one thing which has value in itself, namely the rational or good will that confers value on things. All other things have conditional or extrinsic value. But something can be extrinsically valuable and nevertheless be valued for its own sake. This value theory opens up the route to dealing more satisfactorily with issues of commodification, but it also opens up new ways of dealing with problems of ecological ethics. We can assume that certain things (for example, nature) are valuable for their own sake without having to attribute to them “intrinsic value.” Could you elaborate a bit more on these consequences?

C.M.K.: There are some ways in which the version of Kant’s value theory I advocate is friendly to environmental and ecological issues. In *The Sources of Normativity* I make an argument that we have duties *to* animals and possibly even duties *with respect to* plants. Kant says that we have duties *with respect to* animals, but not *to* animals. I think that might be true of plants, but we can establish a direct duty to animals because of the conscious nature that we share with them. That is something that comes from the way I argue from the Formula of Humanity, the importance in my thought of sharing a value with those you see as having a common identity with you. Being able to value things that you see as belonging to your identity as an animal should make you see that you have an obligation to other animals who have the same kind of experiences. When you realize, for instance, that the disvalue you place on pain and terror springs not from your rational nature but from your animal nature, then this should give you a reason for attributing disvalue to the pain and terror experienced by your fellow creatures.

It is also true, as you suggest, that my account leaves room for saying that objects in nature and beautiful environments are valuable as ends and not just as means even though they are extrinsically valuable. They are extrinsically valuable in the sense that their value depends on their relation to human beings and other animal creatures. That makes room for an argument that is friendly to environmental ethics, although there is a limitation on that argument that I know that some proponents of environmental ethics would not like very much. Inanimate nature in my view is still only valuable insofar as it is valuable *to* human beings and the other animals. It does not have an intrinsic value of its own. There would not be implications for protecting the beauty of the environment

for its own sake. There would, however, certainly be implications for protecting the beauty of the environment for the sake of human beings as valued as ends, but also—when you add that we have duties to the other animals—there would be arguments about protecting the beauty and habitability of the environment for the sake of our fellow creatures.

H.P.-S.: There is, beginning with Schopenhauer's objections to Kant's moral theory, a long-standing criticism that Kant's conception of pure reason amounts to an *ought done for the ought's sake* (ein Sollen um des Sollens willen). This criticism has also been taken up by Anglo-Saxon philosophers, for example Philippa Foot, who argue that "ought" needs a reference point, a connection to an end, to be plausible at all. This line of thought amounts to conceiving of morality as a system of hypothetical imperatives: x ought to be done as x is the means for reaching or realizing y (and y is something valuable). So it becomes understandable why x ought to be done.

The obvious question this raises is that the reference point, that is y, also needs a justification and so we would inevitably have to make use of an appeal to reason per se in justifying y if we want to avoid a regress. Still, the defenders of a hypothetical imperative account might argue that this picture of morality makes better sense. What do you think of this position?

C.M.K.: Well, to get back to something I said before, the position you describe has affinities with consequentialism in the sense that it focuses on acts rather than on actions. The idea of an *ought done for the ought's sake* in this bad sense arises if we treat obligation as a special kind of *purpose* we have in performing acts rather than, what I think it really is in Kant's philosophy, a ground on which we choose whole actions along with their purposes. To say that you are obligated is not to say that duty is your purpose but to say that you see the whole action as one that is necessary for you because of some claim that your humanity or the humanity of somebody else makes on you. So it is not, for instance, that you help this man "because it is your duty" rather than "because he is in need of help." It is rather than you see yourself as obligated "to-help-this-man-because-he-is-in-need-of-help." His humanity calls for that, for you to treat his need as a reason. Obligation is not a substitute purpose but a reason for adopting the whole action including its purpose. The idea that obligation can only function as a sort of substitute purpose arises from the focus on acts; whereas if we take the view I associated with Aristotle and Kant, we also get a different view of where the motive of obligation enters into the picture.

I also think it is an important criticism of the view that morality could be a system of hypothetical imperatives that the hypothetical imperative, as I have

argued,⁶ cannot in any case exist without a categorical imperative. Instrumental thinking can exist in the absence of a categorical imperative, but there could not be a normative instrumental principle in the absence of a categorical imperative. A hypothetical imperative cannot bind us unless the end to which it directs us also binds us. So a system of hypothetical imperatives by itself is a system with no normativity in it anywhere.

H.P.-s.: You think that one of the grave misunderstandings of Kant's theory is that Kant uses a "combat model of the soul," that is, that reason and inclination are two opposing forces within the person and that a person should choose to follow reason. Instead we should see Kant as adhering to a "constitutional model" of the person, namely that a person identifies not with her reason, but with her constitution, whereas this includes reason as well as inclinations.⁷

This seems to be in contrast with the account of inclinations Kant offers in the *Groundwork*. Do you think that this account is overruled by later works, for example *Religion within the Limits of Reason Alone*, and that it should not be taken at face value?

Could you explain how the adoption of the constitutional model is important for giving an account of moral action?

c.m.k.: I do think that this interpretation is in contrast with the account of inclinations that Kant offers in the *Groundwork*, and that Kant came to change his mind about the issue. There is a passage in *Religion within the Limits of Reason Alone* in which Kant says that inclinations are good and to want to extirpate them would be wrong. That remark seems to have as its target the Kant in the *Groundwork*, who wrote that rational beings would prefer to be rid of their inclinations. I associate the constitutional model with Plato as well as with Kant, and it is an interesting fact that the kind of ambivalence about inclinations that shows up in Kant also shows up in Plato. What I mean is this: Plato, like Aristotle and Kant, focuses on the fact that there is both an active and a passive side to our moral psychology; the role of a constitution is to regulate how these things interact in the production of an action. Plato seems at certain moments to think it is somehow regrettable that we have to have a passive side at all. And there are also moments in Kant's texts when he seems to think we have this dream of being like God, pure activity with no passivity at all. It is a funny kind of ambivalence and I am not quite sure what to make of it.

I take the importance of the constitutional model to rest in the account of action it makes possible. On the combat model, reason and passion are regarded as two forces in the soul, and an action is caused by one or the other of them. This doesn't give us an intelligible account of action. A bodily movement

caused by a desire is not an action. And a bodily movement caused by a reason is also not an action, if a reason is just a force essentially like a desire but more intelligent. An action has to be authored by a person, has to come not just from some force within the person but from the whole person. So in order to give an account of what action is, we need a notion of the whole person as the agent, and I think that this is what the constitutional model gives us. It is a model that explains how the person works together as a whole, to be the author of her actions. According to the combat model, a special causal route through the person, so to speak, is what makes an action different from an event: but in that case the action is still just something that happens in or to the person, not something we can attribute to the person as its author. So we lose hold of the objects moral philosophy is supposed to be about—actions and persons considered as the authors of actions—if we accept the combat model.

NOTES

1. Christine M. Korsgaard. "Rawls and Kant: On the Primacy of the Practical." In Hoke Robinson (ed.), *Proceedings of the Eighth International Kant Congress*, Memphis 1995, Vol. 1, Part 3, pp. 1165–73.

2. Christine M. Korsgaard. "From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action." In *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, edited by Stephen Engstrom and Jennifer Whiting, pp. 203–36. New York: Cambridge University Press, 1996.

3. Christine M. Korsgaard. "The Right to Lie: Kant on Dealing with Evil." In *Creating the Kingdom of Ends*, pp. 133–58. New York: Cambridge University Press, 1996.

4. Christine M. Korsgaard. "Taking the Law into Our Own Hands: Kant on the Right to Revolution." In *Reclaiming the History of Ethics: Essays for John Rawls*. New York: Cambridge University Press, 1997.

5. Christine M. Korsgaard. "Skepticism about Practical Reason." In *Creating the Kingdom of Ends*, pp. 311–34. Cambridge, U.K.: Cambridge University Press, 1996.

6. Christine M. Korsgaard. "The Normativity of Instrumental Reason." In *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Gaut, pp. 213–54. Oxford: Oxford University Press, 1997.

7. Christine M. Korsgaard. "Self-Constitution in the Ethics of Plato and Kant." In *The Journal of Ethics* 3, no. 1 (1999).