The Constitution of Agency

Essays on Practical Reason and Moral Psychology

Christine M. Korsgaard

OXFORD
UNIVERSITY PRESS

Contents

Acknowledgments viii
Abbreviations for Frequently Cited Works ix

Introduction 1

Part 1: The Principles of Practical Reason

1. The Normativity of Instrumental Reason 27
2. The Myth of Egoism 69
3. Self-Constitution in the Ethics of Plato and Kant 100

Part 2: Moral Virtue and Moral Psychology

4. Aristotle's Function Argument 129
5. Aristotle on Function and Virtue 151
6. From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action 174
7. Acting for a Reason 207

Part 3: Other Reflections

8. Taking the Law into Our Own Hands: Kant on the Right to Revolution 233
9. The General Point of View: Love and Moral Approval in Hume's Ethics 263
10. Realism and Constructivism in Twentieth-Century Moral Philosophy 302

Bibliography 327
Sources 333
Index 335

Introduction

What constitutes an agent? I believe that we—that is, we rational beings—constitute ourselves as agents, by choosing our actions in accordance with the principles of practical reason, especially moral principles.[1] It sounds paradoxical, I know. How can we constitute ourselves, or choose our actions one way or another, unless we are already agents? How can we take control of our movements, unless we are already in control of them? In the essays in this book, completed with one exception between 1993 and 2003, I develop the Kantian conceptions of practical reason and agency that have led me to this view, and I try to explain how it works.[2] I also sketch and defend an Aristotelian account of the role of our passions, reactions, and emotions in action that I believe coheres well with these Kantian conceptions. And, in Part 3, I discuss some related issues in moral philosophy and philosophical methodology.

The essays are reprinted here with only minor changes, to ensure consistency in style, and in the translations of philosophical classics that I cite. For these essays, while they are primarily constructive rather than interpretive, work with and from the classics of the history of philosophy. In them, I try to think about agency, rationality, and virtue, in the company of Plato, Aristotle, Kant, and Hume, in effect asking them what they think about these issues, and trying to work with the answers that they give. This way of working reflects my deep conviction that the way to make progress in philosophy is to build on the achievements of our predecessors. I do not mean by treating their works as authoritative sources of the truth, of course, but rather by engaging with them in the confidence that real illumination on these topics is there to be found. Where the apparently different views of these philosophers, once properly understood, prove to embody strikingly similar insights—and I believe that this happens far more often than most philosophers suppose—I think we've found as good a place as we could possibly find to go digging for the truth.[3] Where their differences appear to be deep and genuine, we cannot do better than to try to discover and articulate the sources of those differences.

---

[1] This idea is also explored in my book *Self-Constitution: Agency, Identity, and Integrity* (Oxford University Press, 2009). One of the essays in this collection, "Self-Constitution in the Ethics of Plato and Kant" (Essay 3) provides a very short version of the ideas developed in the book.

[2] The one exception is "Aristotle on Function and Virtue," first published in 1986. I planned "Aristotle's Function Argument" as a companion piece to this essay, but never published the earlier version that I wrote of the latter.

In this introduction, I will explain the main ideas I argue for in the three parts of this book.

1. The Principles of Practical Reason

1.1 *Reason and Rationality*

The essays in the first part of the book are devoted to the principles of practical reason. Before discussing the more specific conclusions I reach in them, it will be helpful to say what I mean by "reason."

When we talk about reason, we seem to have three different things in mind. In the philosophical tradition, Reason—I'll use the capitalized form to refer to the general faculty of Reason—refers to the active rather than the passive or receptive aspect of the mind. Reason in this sense is opposed to perception, sensation, and perhaps emotion, which are forms of, or at least involve, passivity or receptivity. Reason has also traditionally been identified with either the employment of, or simply conformity to, certain principles, rational principles, which may include the rules of logical inference, the principles that Kant identified as principles of the understanding, canons for the assessment of evidence, mathematical principles, and the principles of practical reason. A person is called "reasonable" or "rational" when her beliefs and actions conform to the dictates of those principles, or when she consciously and deliberately guides her thoughts and actions by them. And then finally, there are the particular, substantive, considerations, counting in favor of belief or action, that we call "reasons."[4]

What are the relations among these three things? I suppose one might think that they, or some of them, are completely separate things, which have related names more or less by accident.[5] To me it seems more natural to see them as aspects of a single human capacity, and so to relate them somehow, but how? According to one theory, the primary item here is the third thing I mentioned, the reason, a substantive consideration that counts in favor of something—some belief, action, or attitude—and that has normative force.

---

[3] See, in particular, "Self-Constitution in the Ethics of Plato and Kant" (Essay 3 in this volume) and "From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action" (Essay 6 in this volume), for essays that most strongly represent this conviction.

[4] This paragraph is more or less lifted from Essay 7, "Acting for a Reason," pp. 207–8.

[5] John Broome is one example of a philosopher who doubts whether these notions are connected. See, for example, his paper "Does Rationality Give Us Reasons?"

In the case of reasons for action, for example, the fact that an action will bring you pleasure is a reason to do it; the fact that it will harm another person is a reason not to do it; the fact that you promised to do it is a reason to do it, and so on. The principles of reason are simply identified as principles that direct us to act on those considerations, telling us what to count in favor of what. Perhaps they also tell us how to weigh and balance reasons against one another, or in some way how to adjudicate between them when they conflict. And then we call a person "rational" or "reasonable"—that is, we ascribe the faculty of Reason to her—in virtue of the fact that she recognizes and responds appropriately to reasons.

In my own view, there are two related problems with this conception of the relations among the various aspects of reason. First, on this conception the substantive reasons come first, so we cannot appeal to the nature of Reason or to the principles of rationality to help us to identify the substantive reasons. How then are we to identify them, except possibly through the use of intuition?[6] And that brings me to my second objection, which is that this conception does not do justice to the idea that Reason is the active dimension of the mind. Rather, those who favor it envision Reason as a receptive faculty that functions something like a sense, except that what it senses is normative rather than empirical facts.

In the Kantian conception of rationality that I favor, the order of the three aspects of reason goes the other way. Reason—the faculty of reason—is identified first, as the active dimension of the mind, and rational principles are identified as those that describe or constitute rational activity. When those principles are applied to facts and cases, they pick out the substantive considerations that we then regard as reasons.

Taking it more slowly:

The source of Reason is a particular form of self-consciousness that characterizes the human mind. Human beings are conscious of the potential grounds of our beliefs and actions *as potential grounds*. Let me explain what I mean by this. Any conscious animal is guided through her environment by means of her perceptions and her desires or instinctive impulses. Her perceptions constitute her representation of her environment and her desires and instinctive impulses tell her what to do in response to what she finds there. Indeed I believe that for the other animals perceptual representation and desire are not strictly separate. Either through original instinct or as a result of learning, an animal represents the world to herself as a world that is, as we might put it, already normatively interpreted, in the sense that she perceives things in terms of her own interests. She lives in a world that consists of things perceived *as* food or prey, *as* danger or predator, *as* potential mate, *as* child: that is to say, as to-be-eaten, to-be-avoided, to-be-mated-with, to-be-cared-for, and so on. These "normatively loaded" perceptions serve as the grounds of her actions—where a ground is a representation that causes the animal to do what she does.

---

[6] Of course some philosophers who think that the substantive reason is the primary item here think that they can be identified without recourse to intuition. For one example, see T. M. Scanlon's discussion in *What We Owe to Each Other*, chapter 1, section 12, pp. 64–72.

The exact way in which these perceptions or representations operate on an animal's mind to produce her actions may, I now believe, differ in ways that can be ranged along a scale, depending on what sort of consciousness the animal has of her own representations. Primitive animals may respond more or less mechanically to these perceptions; more sophisticated animals may operate with something more like concepts or categories of "food" or "predator" or "threat" to which they respond intelligently; and yet more sophisticated animals may even be aware *that* they and their fellows experience, say, desire or fear. These differences affect the degree of control that the animal has, both over herself and, correlatively, over her environment. Exactly how any given kind of animal's representations give rise to his or her actions is a matter to be investigated empirically. But however it may be with the other animals, there is no question that we human beings are aware, not only that we perceive or desire or fear certain things, but also that we are inclined to believe and to act in certain ways on the basis of these perceptions or desires or fears. We are aware not only of our representations and desires as such but also of the way in which they tend to operate on us. That is what I mean by saying that we are aware of the potential grounds of our beliefs and actions *as potential grounds*.

And this awareness is the source of Reason. For once we are aware that we are inclined to believe on the ground of a certain perception, or to act on the ground of a certain desire, we find ourselves faced with a decision, namely, whether we should do that—whether we should draw the conclusion, or perform the action, on the ground in question, or not. Once the space of awareness—of reflective distance, as I like to call it—opens up between the potential ground of a belief and the belief itself, or between the potential ground of an action and the action itself, we must step across that distance with some awareness that we are doing so, and so must be able to endorse the operation of that ground as the basis for what we believe or do. And a ground of belief or action whose operation on us as a ground is one that we can endorse is a reason. This means that the space of reflective distance presents us with both the possibility and the necessity of exerting a kind of control over our beliefs and actions that the other animals probably do not have. We are active, self-directing, with respect to our beliefs and actions to a greater extent than they are. And it is the same fact that we now both can have, and absolutely require, *reasons* to believe and act as we do.[7]

Where are we to find these reasons? How are we to determine whether our perceptions and desires are adequate grounds for the beliefs and actions to which they incline us? To identify reasons we need principles, principles that we can apply to facts and cases in order to decide whether our impulses to believe and to act count as reasons or not. But as the philosophical tradition shows us, there are many contenders to serve as our rational principles. And this would seem to set us off on a regress. For it appears that we need a reason to conform to one proposed principle rather than another, and, if that is so, there must be a further principle behind every principle, to give us a reason for conforming to it. However—to anticipate my conclusion—there need be no such regress if there are principles that are *constitutive* of the very rational activities that we are trying to perform when we take control of our beliefs and of our actions, in the way that rationality requires of us.[8], [9]

1.2 *Rational Principles*

In the tradition of moral philosophy, three kinds of principles have been proposed as requirements of practical reason.

First, there is the principle of instrumental reason. According to this principle, practical rationality requires us to take the means to our ends. Here there is little dispute about how to formulate the requirement, except that some philosophers regard the ends in question as things desired, while others, such as Kant, argue that a rational requirement can apply only to things willed.

Second, there are versions of what I will call the principle of prudence or rational self-interest, usually understood to require that we maximize the satisfaction of our own desires or interests over time, or something along those lines. It is difficult to give an uncontroversial formulation of this principle, because here there are many disputes. Some philosophers think we are required to maximize the satisfaction only of the desires we have in the present. Others think we must take future desires into account but may discount for the fact

---

[7] This account of the nature of reason is taken with some modifications from *The Sources of Normativity*, especially 3.2.1, pp. 92–4, and "Fellow Creatures: Kantian Ethics and Our Duties to Animals," pp. 85–7.

[8] The ancestor of this argument is to be found in "Morality as Freedom" (CKE essay 6), pp. 164–7.

[9] For another version of constitutivism, see the work of David Velleman, found primarily in his books *Practical Reflection*, *The Possibility of Practical Reason*, and *Self to Self*. Velleman focuses on the idea that action has a constitutive aim, rather than on the idea that it has a constitutive principle. A more crucial difference between us is that he usually identifies that aim as self-knowledge, whereas when I think in terms of an aim, I identify it as autonomy.

So what we say about this case depends on our attitude about the principle of prudence. If we suppose prudence is a rational requirement, we will say: fear prevents Howard from pursuing the end he *ought* to prefer, his overall good, and therefore he is acting irrationally. But if we reject the claim that prudence is a rational requirement, we will say: fear determines what Howard's preferred end is, but there is no irrationality in the case, for reason has nothing to say about which ends we should prefer.

Does Hume think that the instrumental principle, unlike the principle of prudence, is a rational requirement? If he does, then as the argument above shows, there should be cases in which Hume would be prepared to identify someone's conduct as "instrumentally irrational," that is, cases in which, without miscalculating or making a mistake, people fail or decline to take the means to their own "acknowledg'd" ends. Now Hume does not discuss this kind of case, but he does explicitly allow that actions can be irrational in two *derivative* ways: we act "irrationally" when our passions are provoked by non-existent objects, or when we act on the basis of false causal judgments (T 2.3.3,416). Both of these are cases of mistake; the actions that result are not, strictly speaking, irrational. And after discussing them, Hume asserts:

> The moment we perceive the falsehood of any supposition, or the insufficiency of any means our passions yield to our reason without any opposition. (T 2.3.3,416)

This suggests that Hume thinks no one is ever guilty of violating the instrumental principle. Making a mistake, after all, is not a way of being irrational, and Hume thinks we do take the means to our ends as soon as mistakes are out of the way. But this is worrisome. How can there be rational action, in any sense, if there is no irrational action? How can there be an imperative that no one ever actually violates?

The problem is exacerbated when we see that Hume's view is not just that people don't *in fact* ever violate the instrumental principle. He is actually committed to the view that people *cannot* violate it. To see this, we need only consider why Hume might be led to deny that people are ever instrumentally irrational. Offhand, that denial doesn't seem very plausible. People fail to take the means to what they *say* are their ends all the time. And this does not happen only when those ends are demanded by abstract or distant considerations of what will conduce to the person's overall good. It happens in the case of more local ends that are expressly and directly wanted or chosen for their own sakes. You want to ride on this immense roller coaster but you are prevented by terror. Every night of the carnival you go and look at it, get in line for a ticket, and then lose your nerve and shuffle meekly away. You don't think riding the roller coaster is essential to your overall good. Maybe you even think it's risky that they are future. Most agree that any "pure" preference for the present over the future (any preference not based on extraneous factors like the greater uncertainty that attaches to future events) is irrational, but differ about what kinds of items the principle must take into account: all desires, all reasons? The common element in these views is that there is *some* principle requiring us to take the effects on our other ends into account when we reason about how to realize any particular end.

Third, many philosophers have believed that moral requirements are requirements of practical reason. Here, the main distinction is between philosophers who think that the basic moral requirement is formal, like a universalizability principle, and those who think that certain substantive moral principles, like the prescriptions that we should tell the truth and keep our promises, are self-evident rational requirements.

In the first three essays in the book, I take up these three kinds of principles in turn, asking in virtue of what the proposed type of principle is normative—that is, binding upon us—and thereby what qualifies it to count as a rational principle.

In "The Normativity of Instrumental Reason," and (more implicitly) in "The Myth of Egoism," I contrast three possible accounts of the normativity of practical principles generally. According to an empiricist account, the normativity of the principles of practical reason rests primarily in the capacity of those principles to motivate us—in their effects on the will. On this account, the principle of instrumental reason is normative (or, perhaps, does not need to be normative) because we are reliably motivated to take the means to our ends once we know what those are, and the principle of prudence is normative because we reliably prefer the action that leads to our greater good once we see clearly that it does so. The role of reason in action, on this view, is not strictly practical: it is only to clear up mistakes. According to a rationalist or realist account, by contrast, the normativity of practical principles is not something that can be further explained. Certain principles or reasons simply have normative force, as a kind of property. According to this view, reason is supposed to be practical. But, as I mentioned at the beginning, to say that we are practically rational is just to say that we recognize and respond to these normative requirements, in essentially the same way that (according to this theory) we respond to theoretical reasons.

In "The Normativity of Instrumental Reason," I criticize these two accounts. I argue that the empiricist account, while it may explain how we are motivated by rational principles, cannot explain how we can be guided by them, or more generally how they can bind us. A principle that moves us inevitably cannot serve as a guide, for it is not possible to be guided unless it is also possible to fail to be guided. (For these arguments see especially "The Normativity of Instrumental Reason," section 2; and "The Myth of Egoism," section 1.1.) The rationalist account, by contrast, cannot explain why rational principles necessarily motivate us. So long as bindingness or normativity is conceived of as a fact external to the will, and therefore external to the person, it seems possible to conceive of a person who is indifferent to it. But this throws doubt on whether such principles can be binding after all. For what is amiss with a person who is indifferent to his reasons and obligations? He fails to *apply* certain principles to his actions, but then why should he do so? We cannot say that he has a reason to act on his reasons, or an obligation to meet his obligations, without manifest circularity. We can *say* that what is amiss with such a person is that he is irrational, of course, but according to the rationalist theory, that is just to repeat that he does not respond appropriately to reasons. (For these arguments, see "The Normativity of Instrumental Reason," section 3; "The Myth of Egoism, section 1.5; and also "Acting for a Reason," section 3, and "Realism and Constructivism in Twentieth-Century Moral Philosophy," section 4.) If we are to explain the normative force of the principles of practical reason, we cannot just regard them as principles that we are free to apply or not. Instead, like the rules of logical inference, they must be principles *in accordance* with which we operate—either well or badly.

So in place of these unsatisfactory conceptions, I offer a different kind of account of the normativity of the principles of practical reason, according to which the principles of practical reason are constitutive principles of action. I explain how this works in the case of the instrumental principle in "The Normativity of Instrumental Reason," and in the case of the formal principles of morality championed by Plato and Kant in "Self-Constitution in the Ethics of Plato and Kant."[10] I will describe this view of instrumental and moral reasoning more generally before returning to the vexed fate of the principle of prudence or rational self-interest.

1.3 *Constitutive Principles*

But before I go on I must say what I mean by a constitutive principle. First, what I will here call a constitutive *standard* (in the essays, I sometimes use "internal" for "constitutive") is one that arises from the very nature of the object or activity to which it applies. It belongs to the nature of the object or activity that it both ought to meet, and in a sense is trying to meet, that standard. Constitutive standards apply most obviously to objects that have some standard use or function or purpose. If it is the function of a house to provide shelter from the weather, then it is a constitutive standard for houses that they should be waterproof. If it is the function of an encyclopedia to provide information to those who consult it, then it is a constitutive standard for encyclopedias that their statements should be true. Constitutive standards are opposed to external standards, which mention desiderata for an object that are not essential to its being the kind of thing that it is. Of course there is often room for contention about which desiderata are essential, but generally speaking one might suppose it is an external standard for a house that it should have a swimming pool or for an encyclopedia that it should be written in elegant prose.

---

[10] Plato's account of justice in the *Republic* is formal because he regards justice as that principle that unifies or harmonizes the soul, whatever it might be. See R 443e–444a and "Self-Constitution in the Ethics of Plato and Kant," Essay 3 of this volume, pp. 119–20.

Two things are important to notice about standards of this kind. First of all, constitutive standards are at once normative and descriptive. They are descriptive because an object must meet them, or at least aspire to meet them, in order to be what it is. And they are normative because an object to which they apply can fail to meet them, at least to some extent, and is subject to criticism if it does not. This double nature finds expression in the fact that we can criticize such objects either by saying that they are poor objects of their kind ("That's a poor encyclopedia, it isn't up to date."), or by saying that they are not such objects at all ("That's not an encyclopedia: it's just a compendium of nineteenth-century opinion!"). Second, constitutive standards meet challenges to their normativity with ease: someone who asks why a house should have to be waterproof, or an encyclopedia should record the truth, shows that he just doesn't understand what these objects are for, and therefore, since they are functional objects, what they are.

An especially important instance of the constitutive standard is what I will call the constitutive *principle*, a constitutive standard applying to an activity. In the case of essentially goal-directed activities, constitutive principles arise from the constitutive standards of the goals to which they are directed. A house-builder is, as such, trying to build an edifice that will keep the rain and weather out; the writer of an encyclopedia article is, as such, trying to convey the truth. But all activities—as opposed to mere sequences of events or processes—are, by their nature, directed, self-guided, by those who engage in them, even if they are not directed or guided with reference to external goals. And the principles that describe the way in which an agent engaged in an activity directs or guides himself are the constitutive principles for that activity. So it is a constitutive principle of walking that you put one foot in front of the other, a constitutive principle of swimming that you make movements that will impel you forward through the water, a constitutive principle of intelligible linguistic expression that your sentences include both a subject and a verb, a constitutive principle of typing that you hit the letters that you wish to appear on the page, and so on. And in all these cases, we can say that unless you are following the principle in question, you are not performing that activity at all.

Constitutive principles, like constitutive standards more generally, are normative and descriptive at the same time. They are normative, because in performing the activities of which they are the principles, we are guided by them, and yet we can fail to conform to them. But they are also descriptive, because they describe the activities we perform when we are guided by them. Sometimes people are puzzled by the idea that you can fail to conform to a constitutive principle—if following the principle is constitutive of the activity, and you fail to conform to it, then aren't you failing to engage in the activity after all? In one sense that is right, but in another it cannot be, for if you were not engaging in the activity after all, then your failure to conform to its constitutive principle would not be a failure at all. If I am not swimming, but just cooling myself by splashing about in the water, then my failure to make headway through the water is no failure at all. But if I am trying to swim—suppose there is a shark headed towards me—and all I succeed in doing is splashing around in the water, then my failure to make headway is a failure indeed. And this sort of thing does happen—people trying to walk can trip over their feet, people trying to type can hit the wrong letter, and people trying to write can fail to make themselves intelligible for want of a verb. And again, the double nature of the constitutive principle here is reflected in the language we use to describe failures to conform to them. As I watch your inefficient flailing about in the water I can say, with equal force, and meaning the same thing, "You're swimming very poorly" or "You aren't swimming, you're just splashing around." One way to put the point I am trying to make here is to say that the correct account of the metaphysics of activities is a Platonic one. An activity is the activity that it is by virtue of its imperfect *participation* in the perfect Platonic *form* of that activity.

The principles of practical reason, I propose, are constitutive principles of rational activity: they are the principles by which we take control of our beliefs and actions. Or rather, since these terms may already be taken to imply control, perhaps I should say that they are the principles by which we take control of our representations or conceptions of the world, and of our own movements—using "movement" as a general term for the various ways, physical and mental, that we bring about states of affairs in the world.[11]

---

[11] I think that some of the disagreement about whether non-human animals have beliefs or count as agents results from the fact that they have different, and lesser, kinds of control over their

Before I go on, let me notice one complication that arises from my view of reason. I have characterized reason *in general* as a faculty by virtue of which we are active. Of course many philosophers would disagree with that. Some philosophers think that arriving at beliefs is somehow a more passive process than arriving at decisions to act. They think that we cannot help believing the evidence of our senses or the conclusions of our theoretical arguments in some way that we always can help doing what we decide we have most reason to do. Other philosophers, in particular the ones I mentioned earlier, who think that rationality is a matter of responding appropriately to substantive reasons, might deny that there is any such difference, but they would deny it because they think practical reasoning is no *more* active than theoretical reasoning. Both are just a form of responsiveness to the reasons that are there. I disagree with both of those camps, but I will not attempt to engage these large issues here. Since I believe that reason is essentially an active faculty, I regard "action" in the special sense relevant to practical reason as one among several forms of "rational activity." As I am thinking of it, "acting" in the sense relevant to practical reason is that activity that is directed to producing some state of affairs in the world.[12] So action is taking control of the movements by means of which we produce states of affairs in the world. Obviously, any adequate account of action, in this perhaps narrower sense, should also capture the core idea of being active that is common to all forms of rational activity. But for now I will focus on action in this more specific sense. The question, what the principles of practical reason are, is then the question what the constitutive principles of action are: what counts as taking control of our own movements?

1.4 *Agency*

Many of the problems that are now discussed under the rubric of "the philosophy of action" were once discussed under the rubric of "freedom of the will," and this is no accident. Agency is almost as mysterious as freedom of the will, and for the same reasons—with this important difference: that it

---

representations and movements than human beings do. One can grant that, and still think there is no argument worth having about how much control entitles us to call a representation a belief or a representation-directed movement an action.

[12] Of course someone who tries to reason his way to the truth is also trying to bring about a state of affairs in the world—that he has true beliefs about the subject at hand—and to that extent he is acting. But I do not think this thought captures the full sense in which belief is an active state, for this much is common to reasoning your way to true beliefs and manipulating yourself into them, and those are different. Theoretical reasoning has rules of its own. As if probably clear from the text, I am a bit puzzled about how exactly to characterize the relation between action and rational activity more generally.

is much harder for skeptics, even those with "scientific" pretensions, to deny that agency exists. Since I take an action to be a movement that is attributable to an agent, I take agency to be the central notion in the philosophy of action. In virtue of what, then, is a movement attributable to an agent? When can we say that an agent has determined her own movements, and so that those movements are actions? We want to say that a movement is attributable to an agent if the agent is its cause, but this may seem, at first blush, to be in tension with the belief that every event is caused by some other event. How can an agent determine her own movements, if her movements are determined by certain events, which in turn are determined by other events, and so on?

Part of the answer is that there is surely a difference between a case in which the event most immediately determining your movements is, say, that you are pushed from behind, and a case in which the event most immediately determining your movements is a thought of your own. To take the most obvious case: most people do not feel that their freedom or power of self-determination is threatened by the possibility that their movements are determined by their own thoughts about what they ought to do. Rather, they feel that their freedom or power of self-determination is threatened by the possibility that this may *not* be the case. So perhaps we should claim that we are active to the extent that our movements are caused by our conceptions of what we ought to do.[13]

Now I can imagine two possible and opposed reactions to this claim. On the positive side you might react by thinking that it is intuitively plausible: how could we be more in control of our own movements than we are when they are caused by our very own reflections about what we ought to do? But on the more skeptical side you might want to argue that, even granting this kind of mental causation, there is no reason to suppose that thoughts with one sort of content—thoughts about what we ought to do—cause movements that are any more "self-determined" than thoughts with any other sort of content. After all, why should the *content* of the thought make any difference to the degree to which the person moved by that thought counts as a self-determining agent?

Kant's theory of autonomy, I believe, addresses this problem. The will, Kant famously argues, is a kind of causality, and as such, it must operate in accordance with laws.[14] A free will—a fully self-determining will—would be one that is not moved by any *alien* cause. That is, it would not be subject to determination by any law that is outside of itself. Since a free will must operate in accordance with laws, and yet must not be determined by any law outside of itself, the free will must be determined by a law that it gives to itself—a law that it legislates to govern its own movements. The free will, that is, must be an autonomous will. In other words, to be free is to be motivated by the thought that the principle in accordance with which you propose to act is one that you would will as a law. But of course Kant also believed that the moral law is the law of acting on a maxim that you yourself, on your own deepest reflection, would will to be a law—either one that *qualifies* to be a law (when the action is permissible), or one that you *must* will as a law (when the action is required). This means that in Kant's theory autonomy is linked, on the one hand, to the very idea of action—that is, of self-determination—and, on the other hand, to thoughts about what we ought to do. According to Kant, then, to think thoughts about what you ought to do is at the same time to think thoughts about what you would do were you a fully self-determining being.[15] And if it is possible for us to act as we would act if we were fully self-determining beings, then we are, for practical purposes, fully self-determining beings (G 4:446–448). This is why the content of the thoughts that move us can make a difference to the degree of self-determination we exhibit when our movements are caused by our thoughts.[16] The categorical imperative, on this view, is not just the principle of morality. It is also the constitutive principle of action.

---

[13] On my view there are actually degrees of activity or agency, and the phrase "conceptions of what we ought to do" is meant to cover all of them, ranging from a non-human animal's instinctive normative perceptions to a reflective human being's explicit practical deliberations. The argument I go on to give in the text, however, most obviously applies to that last thing: the way we are active when we reflect on what to do. Here again I want to take a Platonic line: other forms of action or self-determination count as forms of action or self-determination because of the extent to which they imperfectly participate in the perfect self-determination that is represented by being determined by explicitly practical deliberation.

More precisely, I believe that the principle of governing oneself by universal laws is the constitutive principle of rational activity generally. For the requirement of universalizability governs every aspect of rational thought. To believe on the basis of a rational consideration is to believe on the basis of a consideration that could govern the beliefs of any rational believer, and still be a belief about the public, shared world. To act on the basis of a rational consideration is to act on the basis of a consideration that could govern the choices of any rational chooser, and still be *efficacious* in the public, shared world. This is what I have elsewhere called the "practical contradiction" interpretation of the categorical imperative test.[17] The notion of efficacy brings in the other element of Kant's account of action, the principle of instrumental reason. For if to act is to engage in practical activity that is directed to producing some state of affairs in the world, then the agent must also seek to be efficacious, that is, to work with the natural causal mechanisms that he can use to make things happen in the world. He must use the means. And this means that the maxim or principle on which he proposes to act must serve as a universal *practical* law. It is the universalizability principle in that specific sense—the law of acting only on universal *practical* laws, which is constitutive of action in the more specific sense.

---

[14] For an explanation of the connection between action and laws see "Self-Constitution in the Ethics of Plato and Kant," Essay 3 in this volume, pp. 120–4.

[15] Take that as a very rough reading of Kant's claim that "*the world of understanding contains the ground of the world of sense and so too of its laws*, and is therefore immediately lawgiving with respect to my will" (G 4:454).

[16] I intend this argument to address the question how we can conceive of ourselves as agents. I do not intend it to address the question of the grounds on which we hold one another responsible. I do not think that attributions of responsibility are directly tied to attributions of freedom or agency in that way. For my views on this matter see "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations," CKE essay 7.

Let me put the point another way. To be an agent is to be, at once, autonomous and efficacious—it is to have effects on the world that are determined by yourself. By following the categorical imperative we render ourselves autonomous and by following the principle of instrumental reason, we render ourselves efficacious. So by following these principles we constitute ourselves as agents: that is, we take control of our movements.

1.5 *Self-Constitution*

If the idea of self-constitution still seems paradoxical, it may be helpful to compare the human agent with another sort of agent whose claims to self-constitution are perhaps less assailable: the political state. In "Self-Constitution in the Ethics of Plato and Kant," I follow Plato in using this comparison to tease out the conditions under which a complex entity can, in Plato's words, "achieve anything as a unit" (R 352a). A state is like an individual human being insofar as all of its actions supervene on other, so to speak smaller, events: in the case of the state, on the decisions and actions of various citizens and office-holders. What makes a certain event or set of events count as an action attributable to the state is that the state has a set of deliberative procedures—a constitution—of which these smaller events can be seen as parts. For example, the constitution might specify that the majority vote of certain citizens who are taken to represent other citizens counts as the enactment of a law. The outcomes of following those procedures, the laws and the execution of the laws, are the actions of the state. Thus the function of the constitution of the state is to unify a diverse group of citizens into a single agent, whose movements count as its actions when they are in accordance with its laws. And so the citizens, by adopting these deliberative procedures, can be said to constitute themselves as a unified agent. In the same way, an individual human being constitutes himself as an agent, by adopting a procedure for the making of laws—a procedure that he requires because of the reflective distance that makes it necessary for him to act for reasons, and therefore on principles.

---

[17] See "Kant's Formula of Universal Law" (CKE essay 3), pp. 93–102.

As Plato argued, however, while there is a formal or procedural sense in which *any* deliberative procedure—any constitution—can unify a diverse group of citizens into a state, there is also a substantive sense in which only a certain kind of constitution—a just constitution—can do so successfully. According to Plato, a just constitution must be one in which the class of the wisest citizens, in the state—or reason, in the soul—rules over the other parts. A state that is ruled, not by the wise, but by the soldiers or by the wealthy or by the common people contains the seeds of civil war within it, because none of these groups can be relied on to govern, as the wisest do, for the good of the whole. And if civil war occurs, then the state can no longer act as a unit. So only a just constitution truly unifies the citizenry into a state. And, for the same reason, Plato thought that a soul ruled by the principle of one of its inferior parts—by appetite or spirit—was in danger of losing the unity that is required for "achieving anything as a unit." Only a just person has the integrity that is essential to agency.

A similar distinction can be found in Kant, for we can distinguish between a maxim that is universal in a merely formal or procedural sense—a maxim that an agent may tell himself, perhaps with insufficient reflection, that he is prepared to will as a law—and a maxim that is actually, substantively, universalizable. A substantively universal maxim is one that an agent really could, with all due reflection, will to be in effect, and to govern his own conduct, in all relevantly similar circumstances. Thus, according to both Plato and Kant, just as the agency of the state is constituted by the adoption of deliberative procedures whose perfect realization depends upon political justice, so the agency of the soul, of the human individual, is constituted by the adoption of deliberative procedures whose perfect realization depends upon personal justice or morality. The unity that is essential to agency and moral integrity are one and the same thing.

1.6 *The Problem of Prudence*

I now return to the other proposed rational principle, the principle of prudence or self-interest. In "The Myth of Egoism," I argue that some common assumptions about this principle cannot be right. The principle is not, as many social scientists seem to assume, either identical to, or a mere application of, the principle of instrumental reason. For if there is a rational requirement of prudence, it does not merely require us to take the means to some end that we already or inevitably have. Rather, it requires us to *have* a certain end—one in which our more particular ends are somehow harmoniously combined—and to always prefer that end to all of those more particular ends themselves. The principle of prudence, if it exists, is therefore a principle of pure practical reason. How is such a normative principle then to be established? Part of the difficulty is that the idea of prudence, as usually conceived, is trying to do a double job. Prudence is usually supposed to require us to take into account all of the ends we have reason to promote, including those we will have in the future, whenever we deliberate. But it is also supposed to require us to be (especially? exclusively?) attentive to what is in some difficult-to-define sense our own personal good or interest. I call those two elements of the principle the requirements of balancing and of particularity, respectively. In the essay I argue that no intelligible account can be given of the requirement of particularity (see especially sections 2.1–2.3). More generally, it seems obvious to me that a requirement of self-interested prudence could not possibly be established in the way that I believe the principle of instrumental reason and the categorical imperative can. That is, it could not be a constitutive principle of action that we pursue our own overarching good. If acting is determining yourself to be a *cause* of some state of affairs, then you are *just not acting* unless you take the means to that state of affairs. If acting is determining *yourself* to be a cause of some state of affairs, then you are *just not acting* unless your movements are determined by a law that you give to yourself. But we certainly cannot argue that you are *just not acting* unless you pursue your own overarching good.

However, I believe that perhaps this kind of foundation may be given for the requirement of balancing, taken separately from any thought about particularity or self-interest, although I do not now see my way to an argument to this effect. I believe this partly because the requirement of balancing does seem to me to be a requirement of practical reason, and indeed one without which the theory of practical reason is radically incomplete. And I believe it partly because it seems plausible to say that our unity as agents depends on our conformity to such a requirement. If that is right, then the principle of balancing may also be a constitutive principle of agency, and so a principle of practical reason.

2. Moral Virtue and Moral Psychology

Our agency depends in important ways on the character of our more passive or receptive faculties. Human action partakes of *reaction* at least to this extent: something must make it occur to us that we might perform a certain action. And this means that our ability to do what reason demands also depends on these faculties. In the examples used to illustrate the duty of beneficence in the *Groundwork*, Kant envisions someone to whom it occurs to help, but who doesn't wish to: he tests the maxim of *not* helping, and finds that he cannot will it as a universal law (G 4:423). But suppose that he has so little sympathy that the thought of helping doesn't even come into his mind? A Kantian agent with the most determined resolution to test all of his maxims by the categorical imperative would not succeed in meeting all of the requirements of duty if certain things never occurred to him. And more generally when we are formulating our maxims, with a view to testing them, we need to know what features of our situation are relevant to the question whether we have good reason to act as we propose to do or not. Our attention must be directed in the right ways, and this depends on our receptive faculties—on what we feel and what we notice. Kant himself noticed at least the first of these problems. In *The Metaphysical Principles of Virtue*, after quoting the universal law formulation of the categorical imperative, he remarks:

> Maxims are here regarded as subjective principles which merely *qualify* for a giving of universal law, and the requirement that they so qualify is only a negative principle (not to come into conflict with law as such).—How can there be, beyond this principle, a law for the maxims of actions? (MM 6:389)[18]

By a law for the maxims of actions, Kant means a law making it necessary to have certain maxims. And he takes this problem as the point of entry for his own account of virtue.

The term "virtue" may be used broadly to refer to one's moral condition in general, or more narrowly to refer to the possession of certain dispositions that make one *receptive* to the demands of reason. And for the reasons I have just sketched, these will include dispositions of our desires and emotions, as well as of reason and will. The first three essays in the second part of this book, taken together, sketch a reading of Aristotle's *Nicomachean Ethics*, which I take to be centrally concerned with the question of virtue in this sense. I do not believe that Aristotle's is a "virtue ethics," if that is supposed to mean that he believes that the fact that something would be chosen by the virtuous person, or by the practically wise person, is what *makes* it right. What makes an action right in Aristotle's theory, I think, is that it is both an expression of, and promotes, the successful performance of the human function, which is rational activity. Aristotle thinks that the person of practical wisdom, being most susceptible to the demands of reason, is best able to *identify* the action that is right in this sense. I believe that this account of virtue is available to the Kantian as well. Aristotle certainly gives a greater role to receptivity and a lesser role to ratiocination in identifying the right than Kant does, but these are matters of degree, and do not render the two views incompatible. And if you substitute "autonomy" for "rational activity" in the account of what makes an action right that I just ascribed to Aristotle, their views about what makes actions right look fairly similar as well.

---

[18] Both Mary Gregor and James Ellington (in his translation of the *Metaphysical Principles of Virtue* in *Ethical Philosophy*) translate the parenthetical remark with the phrase "a law"—as in Gregor's "not to come into conflict with a law as such." I have deleted the "a." What Kant means is not that the maxim is not supposed to conflict with some particular law; he means that the maxim is not supposed to conflict with the very idea—the form—of law. See "Kant's Analysis of Obligation: The Argument of Groundwork I" (CKE essay 2), pp. 61–2.

As I've already hinted, my reading of the *Nicomachean Ethics* turns on Aristotle's famous "function argument" in 1.7, and the first of the three essays, "Aristotle's Function Argument," is a defense of that argument. Aristotle argues that in order to ascertain the good for human beings, we must first identify the human function, and he settles on rational activity as the answer (NE 1.7 1097b22–1098a17). There are many evident objections to this argument (which I review and reply to in the essay) and as a result some philosophers have apparently thought that the reader of Aristotle should simply set it aside. But the argument plays an essential structural role in the project of the *Nicomachean Ethics*. The Greek philosophers recognized a conceptual connection between function (*ergon*) and virtue (*arete*). A thing's virtues are the properties that make it good at performing its function. It is an essential part of Aristotle's project to show that the qualities that are ordinarily considered moral virtues are really virtues in this more technical sense, since the desirability of the virtuous life turns on the fact that virtues make us good at performing our function.

That point sets up the project of the second essay in Part 2, "Aristotle on Function and Virtue," which is to ask how exactly having certain emotional reactions *could* contribute to rational activity, and to examine the merits of various possible answers suggested by the text of the *Nicomachean Ethics*. This essay, unlike the others in the volume, is an older essay (1986), and there are many things in it that I would not now say, or would say in different ways. But I still endorse its conclusions, both as an interpretation of Aristotle and, at least to a large extent, as an account of virtue. Aristotle seems to suggest, at different moments, that the emotions contribute to rational activity by giving way to reason, by harmonizing with its demands, by making the agent susceptible to its influence, by promoting rational activity in the way that good physical habits promote physical activity, or by enabling us to perceive what is good. The conclusion I reach combines several of those ideas: that the emotions constitute a kind of *perception* of the good, or of reason, that at once makes us susceptible to reason's influence and helps us to form and act on correct *conceptions* of the good.

Now let me return to the question what all this might have to do with Kant's theory of rational agency. My views about the compatibility of Kant and Aristotle are somewhat unorthodox, and defending them sets the project of the third essay in Part 2, "From Duty and For the Sake of the Noble: Kant and Aristotle on Morally Good Action." Here I argue that Kant and Aristotle share an important view about the locus of moral value, and a related thesis in moral psychology, which sets them apart from most other moral philosophers. In order to describe this view here I will use some technical terminology that I set out most plainly in the fourth essay in Part 2, "Acting for a Reason," especially section 4. The thesis is that moral value does not attach to mere *acts* (to use a couple of Kant's examples: making a lying promise, committing suicide) but rather to what I call *actions*: an act performed for the sake of a certain end (making a lying promise for the sake of personal gain, committing suicide in order to escape your troubles). A Kantian maxim, the sort of thing tested by the categorical imperative test, and an Aristotelian *logos*, its analog, always include (at least) both an act and an end, and it is *these* formulations that describe the locus of moral value, the objects that we deem to be obligatory or forbidden, noble or base. What's forbidden or base is not, for example, telling a lie, but telling a lie for the sake of personal gain. The associated thesis in moral psychology is that these formulations also describe the objects of human choice. According to both Aristotle and Kant, human beings do not just choose acts, motivated by ends that are foisted upon us by natural forces (e.g. desires). Rather, the object of a rational choice is a certain act performed for the sake of a certain end, and we decide whether to do that by asking whether the whole action described by that formulation is a thing worth doing for its own sake. According to Kant, we ask whether a certain act performed for a certain end is required by duty, permissible, or forbidden; according to Aristotle, we ask whether it is noble, not ignoble, or base.

I believe that the main disagreement between these two philosophers, concerning the role of the emotions, does not spring from their moral theories, but rather from their views about the nature of the emotions. Aristotle's theory of the virtues, as I interpret it in "Aristotle on Function and Virtue," is predicated upon his view that pleasure and pain—and therefore the emotions that always involve them—are perceptions of good and bad, or as we might say of reasons. To feel fear, for instance, is to feel oneself in the presence of a reason to flee. But Kant does not believe that pleasure and pain, and with them the emotions, are perceptions: he considers them to be mere feelings, which, as he puts it, express nothing at all in the object but simply a relation to the subject (MM 6:211–212). So when he famously denies, in the first section of the *Groundwork*, that there is any moral worth in action motivated by sympathy (G 4:398), it is important to keep in mind that he is thinking of sympathy, not as an inchoate perception of the fact that there is reason for helping, but rather as something like a *taste* for helping. His sympathetic man *likes* helping, the way one might like chocolate. If we think of the sympathetic man's sympathy instead as an inchoate perception that there is reason for helping, we will disagree with Kant about this example. But this will not make it necessary to disagree with the conclusions Kant draws about the moral law, for what sympathy perceives is a normative demand that is grounded in the other's humanity. On Kant's view of the emotions, they could at most be instrumental aids to moral choice, accidentally pointing us in the right direction, whereas on Aristotle's they are a more intrinsic part of our grasp of our reasons to act.

Is Kant stuck with the view that emotions are mere feelings or tastes? One might suppose that only a substantive realist about reasons could treat them as objects of perception, and so only a realist about reasons could interpret the emotions as perceptions of reasons.[19] But I do not think this is true. Any animal, in perceiving the world, perceives it through the lens of her own cognitive requirements and interests, perceives it in a way that enables that particular animal to find her way around. And as I mentioned earlier, I think it is the nature of every animal to have normative perceptions—to see the world through the lens of her own concerns and interests, or, as we might say, her values. And this is true of us as well. The implication for rational beings is that the development of rationality requires the acquisition of a second nature—a set of emotional responses and an accompanying normative view of the world that conforms to the demands of reason. The acquisition of virtue, a condition of the receptive faculties that makes us sensitive to the demands of reason, is therefore essential to the perfection of our moral nature, and to the integrity that makes agency possible.

In the last essay in Part 2, "Acting for a Reason," I put Aristotle and Kant's view about the locus of moral value and the object of human choice to work, using it to address some contemporary issues about the ontology of reasons and the nature of rational motivation. I start from the debate between those who think that desires and other mental states are reasons for action and those who think that only certain facts—facts about what is good about the action or its effects—can provide us with reasons for doing it. I argue that reasons have an essentially reflexive structure not captured by either of these accounts. The person who acts for a reason must be motivated by certain facts about the action, but he must at the same time be motivated by the awareness that these facts constitute a reason. An account of rational motivation must show how these two elements in an agent's motivational state can be combined. The issue is similar to that presented by the Kantian idea that good moral agents do the right thing from duty. Many readers criticize this idea because they think that acting from duty must be an *alternative* to acting for the sake of some concrete end such as the welfare of others. They think it is intelligible to ask whether someone acted from duty, or because the other person's welfare was his end. No one thinks it is similarly intelligible to ask whether someone acted for a reason, or because the other person's welfare was his end. Yet the problem is the same: to show that these two elements in motivation are not alternatives, but can be combined. Aristotle and Kant's view that value and choice attach to whole "actions" rather than mere "acts" solves this problem, for they show that acting "from duty" or "for the sake of the noble" is not an alternative to acting for the sake of certain ends. Rather, for example, the agent who helps from the motive of duty judges that it is his duty to make the welfare of another his end. More generally, on their theory the choice of an action has the required reflective structure: the maxim or *logos* specifies what is good about the act—say, that the act realizes a certain end—and the agent chooses the whole action on the basis of an assessment of its value as a whole. Only theories of this kind, therefore, can give a satisfactory account of what it means to act for a reason.

---

[19] See "Realism and Constructivism in Twentieth-Century Moral Philosophy," Essay 10 in this volume, for the difference between realism and Kant's constructivist account.

3. Other Reflections

The three essays in Part 3 of the book are admittedly miscellaneous; in these remarks I will focus on their links to the major themes of the book. In "Taking the Law into Our Own Hands: Kant on the Right to Revolution," I take on the problem posed by Kant's paradoxical attitude towards revolution. Kant's personal enthusiasm for the French Revolution earned him the nickname "the Old Jacobin," but in the *Metaphysical Principles of Justice* he argues that political revolution is always wrong (MPJ 6:320). I defend this last conclusion on the grounds that the political state must function as a unified agent, whose constitution makes its government the voice of its general will. Given Kant's view that justice requires that we regard ourselves as citizens of the political state, we must regard ourselves as subject to the general will. But I argue that in certain circumstances a virtuous person might nevertheless be morally motivated to instigate or participate in revolution. Here the distinction between procedural and substantive justice to which I appealed in "Self-Constitution in the Ethics of Plato and Kant" again plays a role. For the argument I've just gestured at depends on a procedural conception of justice, and there can be circumstances in which what is procedurally just is not substantively so. In these cases an individual with the virtue of justice is divided against himself, and must make a judgment about what to do which is not governed by any principle that we share with others and therefore may not be justifiable to them. The case vindicates Plato's view that we cannot be fully unified as individuals if the state in which we live is unjust and therefore is itself disunified.

To the extent that that is true, our efforts at self-constitution may be affected, even limited, by what goes on in the world around us. This is an aspect of the subject about which I have written little. The second essay in Part 3, "The General Point of View: Love and Moral Approval in Hume's Ethics" touches on a view that carries this thought to the opposite extreme, for on Hume's view, our agency is entirely constituted in the eyes of those around us. This conclusion naturally follows from Hume's view that causality in general is something that exists in the eyes of beholders—that is, those who observe the constant conjunction of certain events—together with the fact that to be an agent is to be a cause.

The question of how we conceive of agency in Hume's theory arises obliquely in the essay, which is directed to a question about Hume's doctrine of the general point of view. In Hume's theory moral virtue depends on moral approval rather than the reverse: a moral virtue is, essentially, a quality of which spectators approve. And Hume argues that moral approval, in turn, is a calm form of love we feel when we view someone from the general point of view. To view someone from the general point of view is to view him sympathetically from the point of view of his usual associates, and to consider the general effects of his conduct rather than the specific effects of his actions in this or that case. My own argument starts from the question why we should take up the general point of view in the first place. Given that it is from this point of view that the very idea of virtue arises, it cannot be that we take up the general point of view in order to observe someone's virtues more clearly. If we did not take up the general point of view, we would not see people as having virtues at all. Nor, I argue, does Hume's own account of why we take up the general point of view—in order to reach agreement with others and make consistent judgments within ourselves—work. For again there is no virtue to reach agreement *about* until after we take up the general point of view. On Hume's behalf I construct an argument to the effect that our capacity to view another as a regular cause and so as an agent depends on taking up the general point of view. I trace our motivation for taking up that point of view to the demands of love, which seeks to view its object as a person with a character, and so as an agent. The resulting view gives us an account of why we view the world morally that I think can be motivated by Hume's theory even if it was not his own. It also gives us a picture of how agency might be constituted from the outside in. While I don't share Hume's view that our characters exist primarily in the eyes of beholders, I do believe in a more moderate implication of his view: that sympathy with others shapes and limits our powers of self-constitution in important ways.

In the final essay in the volume, "Realism and Constructivism in Twentieth-Century Moral Philosophy," I turn to more methodological issues. The essay was written as a lecture for (as it turned out) two occasions on which I was asked to speak about the issues and achievements of moral philosophy at the turn of the century. In it I try to describe the difference between a realist and a constructivist moral theory. I have long believed that it is unhelpful to characterize moral realism as a view about whether moral statements are true or false. This makes it seem as if realism and early non-cognitivist theories—views according to which moral statements are the expressions of attitudes rather than truth-apt assertions—are the only available options. That in turn lumps together as "realist" theories that are different in important and systematic ways. In *The Sources of Normativity*, I tried to capture this thought by distinguishing "substantive" from "procedural" realism (SN 1.4.4, pp. 34–7). Here, I argue that the difference between a realist and a constructivist theory rests in the way the two views understand the function of concepts, rather than in their views about the truth-value of sentences.[20] A realist believes that the function of concepts is to describe the world, to mark out the entities we find there, while a constructivist believes that the function of (at least some) concepts is to mark out, in a schematic way, the solution to some problem that we face. The task of the philosopher is then to identify the content of such a concept by working out the solution to the problem, thus providing a particular *conception* of whatever the concept names.

John Rawls's work provides an example of what I have in mind. In his philosophy, the concept of justice schematically names the solution to a problem: "justice" names whatever solves the problem of how we are to distribute the benefits and burdens of social cooperation. We have the concept of justice *because* we face the problem, not because justice is something we have encountered in the world. Rawls's own two principles of justice are a conception of justice, worked out by imagining the way a suitably situated group of people who are actually faced with the task of choosing principles of justice—people in the original position—would reason about what their principles should be.

---

[20] For a remarkably similar view, but in an expressivist context, see Allan Gibbard, *Thinking How to Live*, e.g., pp. 79–82, 102–5. For instance, Gibbard says, "There is no contrast to be drawn between ethical and natural properties. The contrast is between ethical and naturalistic concepts" (p. 105).

Once we have identified the correct or best conception for a given concept, there is no reason to deny that sentences involving the use of that concept can be true, for the concept is applied correctly when it is applied in accordance with the correct or best conception. For a Rawlsian to say that some policy is just, is, unpacked, to say that it is in accord with the principles that would be chosen in the original position and that those principles represent the correct or best solution to the distribution problem. There is nothing there that is not truth-apt, for anyone who disagrees is disagreeing about something real—either about whether the justice of the policy does follow from those principles or about whether those principles represent the correct solution to the problem that gives us the concept of justice. Yet moral philosophy as conceived by constructivism is a practical enterprise, an enterprise of working out the solutions to problems, not a mysteriously non-empirical theoretical enterprise aimed at identifying normative facts that are somehow part of the external world. The difference between substantive realists and constructivists does then not rest in a disagreement about the truth-value of sentences deploying normative terms; rather, it rests in a larger difference of approach to the subject: the constructivist regards moral philosophy as a form of practical thinking all the way down.

I see my own view as constructivist in this sense, and the remarks I made in the first part of this introduction, in particular about how we might reason from the nature of agency to the principles of practical reason, are intended as a constructivist argument. I could perhaps make that structure clearer if I put the argument this way: The reflective distance produced by the awareness of the potential grounds of our actions as grounds confronts us with a problem, a problem we face as rational agents. That is the problem of how we are to exercise our autonomy efficaciously in the world. And the categorical imperative and the principle of instrumental reason, as the laws of autonomy and efficacy, are the solution to that problem. They are therefore the principles by which, with the aid of the virtues, we constitute our agency.

Christine M. Korsgaard
January, 2008