

# The Normative Constitution of Agency

Christine M. Korsgaard

Harvard University

## *Part I: Natural and Normative Conceptions of Agency*

### 1.1. Two Conceptions of Agency

The original aim of this paper was to throw some light on the question whether, when people act together, there is some way in which they are able to share their agency itself, and constitute a single, genuinely unified collective agent, rather than merely a group of separate agents coordinating and cooperating in various ways.<sup>1</sup> But the bulk of the paper is devoted to another question, which I believe represents a deeper issue that is at stake in the question about collective action. This is a question about the status of agency itself, and whether we should conceive of agency as something that is “natural,” or something that is, as I will put it, “normatively constituted,” at least in part. Or rather – since we should not presuppose that we have to choose between these

---

<sup>1</sup> The paper was originally written as a keynote lecture for the *Conference on Collective Intentionality VII: Perspectives on Social Ontology*, held in Basel, Switzerland, in August 2010. I am grateful to the organizers for inviting me, and to the audience on that occasion for their responses. I am also grateful for audience responses and discussion at Wesleyan University and the Goethe University in Frankfurt.

two conceptions – I should say it is a deeper question about what the relationship between these two conceptions of agency is.

When I say that agency is natural, I mean, speaking roughly, that we can identify agency in terms of the particular kind of causal pathway that leads to an action. I have in mind views like those in the Davidsonian tradition, according to which an action is a movement caused or causally guided by certain mental states, such as belief/desire pairs, or intentions, or the desire to act for reasons, say, and the action is attributed to the agent in virtue of the fact that the agent is the one whose mental states brought the action about.<sup>2</sup> The task of the philosophy of action is then to see which mental states can ground the special form of attribution that is involved in the identification of someone's movements as his action. I will come back to the question what is special about that form of attribution later on.

The conception of agency as something that is normatively constituted, on the other hand, is familiar to us from the tradition of political philosophy, where the political state is often considered as a kind of agent, characterized by what Rousseau called a General Will.<sup>3</sup> In such conceptions, the state's capacity to act as

---

<sup>2</sup> The various options listed here are suggested by the work of Davidson himself, Michael Bratman, and David Velleman, respectively (for references, please see the bibliography). They are not intended to be exclusive of each other.

<sup>3</sup> Jean-Jacques Rousseau, *On the Social Contract*. Edited by Donald A. Cress. Indianapolis, Hackett Publishing Company: 1988.

a single unified agent depends upon the establishment of normative relationships among the people who make it up, in particular their relationships of authority and mutual obligation. Speaking roughly, the state can perform an action that counts as its own because someone – some person or body of people – has the authority to make a decision on its behalf, and so to “speak for” the state. But it is important to notice that when I say that agency is “normatively constituted,” I mean both that the *capacity* for agency consists in or depends on the existence of certain normative relations, and that the *realization* of that capacity – success in action – depends on conformity to the norms in question. By success in action, I mean not merely (or even necessarily) the achievement of the agent’s end, but success in actually *performing* an action, in doing something that *counts* as an action. For when we think of agency as something that is normatively constituted, the very idea of an action has a certain honorific character. The agent has to earn the right, we might say, to claim a certain movement or the achievement of a certain effect as *his own*, and what he does only counts as an action if he can so claim it. To use the example I used in my book *Self-Constitution*, an attack by a Spartan on an Athenian only counts as an act of war performed by *Sparta* if the Spartan is a soldier acting under the orders of Sparta’s rulers.<sup>4</sup> When the political state is conceived as an agent, its capacity

---

<sup>4</sup> Korsgaard, *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009, §7.2.11, p. 142.

to act is constituted by the authority relations that determine what counts as a decision made by the state, but its success in action also depends upon conformity to the norms that govern those relations. That is an essential part of what it means to say that the agency of the state is normatively *constituted*.

In contemporary philosophy, the idea that normative relations are constitutive of shared or collective agency has been advanced by Margaret Gilbert.<sup>5</sup> Of course the political state itself is a kind of collective agent, composed of the citizens<sup>6</sup> acting together. So we might suppose that the two conceptions of agency that I have just sketched simply belong to two different forms of agency: individual agency is a completely natural phenomenon, while collective agency is at least in part normatively constituted.

But such an asymmetry seems inherently objectionable. The claims of genuinely unified collective agency to be a real phenomenon must rest on its symmetry with individual agency, or at least the two conceptions of agency must have something recognizable in common. But if normative constitution in the case of collective action is a matter of establishing, and conforming to, normative relationships among the parties to the action, what would it even *mean* for

---

<sup>5</sup> See, for example, Gilbert, "What is it for *Us* to Intend?" in her *Sociality and Responsibility* (Rowman & Littlefield, 2000), and "Shared Intention and Personal Intentions," *Philosophical Studies* 144: 167-87, 2009.

<sup>6</sup> More properly, of people who take on the role of "citizens."

individual agency to be normatively constituted? In particular, how can an individual be normatively related to herself?

In *Self-Constitution*, I discuss two models of individual agency, derived from the philosophical tradition, according to which individual agency is normatively constituted. In Kant's theory, agency is the exercise of autonomy, and autonomy involves the adoption of a maxim to govern one's action, which Kant understands as making a law for oneself. The agent's self-consciousness, in the sense of her awareness of the grounds of or potential reasons for her actions, at once creates the need for a law to govern her action, and places her in a kind of relationship of authority over herself.<sup>7</sup> So there is a sense in which an agent's capacity for action rests in her capacity to make laws that govern her own movements, and her success in action – or so I argue – depends on whether the maxim on which she acts is one that she can actually will, on adequate reflection, as a universal law. If Kant is right, that means that her success in action – here I mean success in actually performing an action, as I discussed before – depends on whether her action is morally good.

The other account I discuss comes from Plato, who models his account of individual agency on his account of the agency of the polis or city-state. Since the doings of the people in a state only count as the actions *of* the state if the

---

<sup>7</sup> See Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), Lecture 3, pp. 90-130; *Self-Constitution*, Chapter 6, pp. 109-32.

proper authority relations obtain between them, Plato's account suggests that the doings of a person only count as his own actions when the proper authority relations obtain among the parts of his soul. Reason must rule for the good of the whole, and appetite and spirit must willingly conform to its rule. Plato calls such a condition "justice" and argues that only someone with a just soul is capable of "accomplishing anything" – that is, as I read it, is really able to act.<sup>8</sup> Again, moral goodness is constitutive of success in action. In both theories, morally good actions are good as actions, good of their kind.

Such views, especially Plato's with its talk of parts of the soul, may sound fanciful to a modern audience, but I believe that there is an important thought at work in these conceptions of agency. These philosophers suppose that the kind

---

<sup>8</sup> Plato does not directly address the question what counts as an action. But he argues that justice is what makes it possible for a state or a soul to achieve anything "as a unit" (*Republic* 352a) which I take to mean that justice is what makes it possible for a city or a soul to give rise to actions and effects which can be identified as its own. He first establishes that even a band of robbers or thieves must be just to each other if they are to achieve anything. This shows that achieving anything requires an internal condition of justice. So an unjust individual, one in whose soul the proper authority relations do not obtain, is incapable of achieving anything (*Republic* 352b). Of course, one might take Plato to mean merely that an individual who is somehow at odds with himself is likely to bumble around and be ineffective. But both the word "incapable" and the addition "as a unit" – "incapable of achieving anything as a unit" suggest a stronger point, and the stronger point is supported by the city/soul comparison at the heart of Plato's argument. Acts that do not issue from the proper procedures – the ones defined by the normative roles of the offices of the state – do not count as the state's own actions at all. Quotations are from the translation of the *Republic* by G. M. A. Grube, revised by C. D. C. Reeve, in *Plato: Complete Works*, edited by John M. Cooper. Indianapolis: Hackett Publishing Company, 1997.

of *unity* that is essential to the notion of agency is something that must be normatively constituted, that is, something that is achieved by conformity to certain norms. What it is to make a decision, on this conception of action, is to place yourself fully behind the movement you are about to make or the change you are about to effect, to endorse it wholeheartedly. Decision takes place against a background threat of disunity – a threat that it is part of the essential function of decision to surmount. In Plato’s political model, the different parts of the soul compete for its rule. In Kant’s account of moral choice, reason and self-love compete for the authority to make the law. In almost anyone’s account of imprudence, present desire uses its unchallenged incumbency to silence representatives of the future self. It does not really matter in which of these terms we conceive the threat; it does not even really matter whether we take the “parts” or factions involved to be functional or temporal. The point is that the complexity introduced into the human psyche by self-consciousness introduces a standing threat of disunity that must be overcome before the agent can act. So there is a threat to individual agency that parallels the standing threat of disunity among the parties who are supposed to be cooperating that is the most obvious problem for collective agency.

Why is disunity a threat? Why is unity essential to agency? Unity is essential to agency, whether collective or individual, because an action, unlike other events whose causes in some way run through an agent, is supposed to be

a movement, or the effecting of a change, that is backed by the agent as a whole. It is a movement that we attribute to *him, to the agent*: not just to his eyebrow like the twitch, or to his feet like the slip, but to the agent himself. Unity, in other words, is necessary to ground the special kind of attribution that is distinctive of agency: attribution to the agent himself.

## 1.2. Implications of the Attribution of Agency: Activity and Identity

Before I discuss that special kind of attribution, I want to talk about why we might *favor* a naturalistic account of individual agency. Although there may be unclear cases, we are not generally at a loss to determine which processes or events are actions, and which of the things around us count as agents. The natural world around us, including the non-human world, plainly contains both. If I walk to the front door in order to leave the house, I am performing an action, and if my cat walks to the front door in order to greet me when I return, she is performing an action too. In both cases something quite different is happening than when an earthquake tips the back of the house upwards and my cat and I tumble helplessly in the direction of the front door. We all readily agree that digestion and circulation are not movements we undertake intentionally and therefore are not actions, and that is true whether they are occurring in me or in my cat. Automatic reactions and expressive phenomena, like ducking or weeping, are not undertaken intentionally, but they may be intentionally

controlled or suppressed, so they seem to fall somewhere in between.<sup>9</sup> Scientists tell us that in dolphins, breathing is under voluntary control in a way in which, in humans, it is not. We humans can suppress breathing, at least for a short time, but we do not have to breathe on purpose, and dolphins do. Half of their brains must always be awake, lest they forget to breathe. So for them, breathing is *doing* something, like walking is for us. If scientists can tell us which sorts of things are under an animal's voluntary control, then surely, the idea of "doing something" is to that extent a natural notion. Which notion? The naturalistic conception of agency – that an agent is active when her movements are caused or causally guided by her own mental states or representations – seems like a natural choice.

But there are two implications of the kind of attribution that we ordinarily take to be distinctive of agency (at least when that notion is applied to human or rational beings) that may be difficult for to explain on a purely naturalistic conception. The first implication looks, on the surface, metaphysical: an agent's *activity* is supposed to be implicated in her agency. Agents are supposed to be efficacious: they *make* things happen. Attributing an action to someone is supposed to be different from attributing either a property or an involuntary movement to her – different from saying, for instance, that she has brown eyes or is tall or that she stumbled and fell – in an important way. It is supposed to

---

<sup>9</sup> They are also *meaningful* or *intelligible* in something like the way action is, and tumbling across the floor during an earthquake is not. But it is not clear where that thought fits in.

indicate that the agent's capacity for being creative or originary in some way, spontaneous in the Kantian sense, has been at work. The efficacy of agency is the agent's own efficacy, we might say, not just that of some chain of causes running through her. While this looks, as I said, as if it is primarily a metaphysical point, it also seems to have a normative implication: An agent, just as such, is the kind of thing that can succeed or fail. Agency is not just a particular form of causality, because causes, just as such, cannot succeed or fail. It is not immediately obvious how this feature of the concept of agency can be captured by an account that explains agency in terms of the causality or causal guidance of an appropriate mental state. The reason is familiar to us from the old debates about freedom of the will – the mental state *itself* presumably has prior causes, so it is not clear why its operation should especially represent the agent's own activity or spontaneity, any more than anything else.

The second implication looks, on the surface, normative: to say that someone did something is to mention what is, in general, a proper ground for responses to her that are normative or at least personal. Characteristically, we hold adult human beings responsible for their actions; more generally, we take people's actions to be legitimate or at least appropriate grounds for responses like love and hate, liking and disliking, gratitude and resentment. In all of these ways, we take a person's actions to reflect something really essential about *her*, to represent her in some way. As I will put it, a person's *identity*, her essential self,

seems to be evinced in some special way in her actions, so that when we respond to her actions, we are responding to *her*. Again, there is a question about how the legitimacy or appropriateness of these reactions can be explained by an account of agency in terms of a causal pathway that leads through the agent's mental states. We can ask why exactly causation by a certain mental state counts as an expression of the agent's essential identity.

So on our ordinary concept of agency, when we attribute an action to a person, we are suggesting that he has been active, and when we respond to someone's actions, we take ourselves to be responding to *him*, to features of his identity.<sup>10</sup> I am going to call those two features of our ordinary concept of agency *the activity implication* and *the identity implication*.

In one way, these two implications seem closely related to each other: both reflect ways in which the idea of an action is supposed to be the idea of a movement, or the effecting of a change, which has its ultimate *source* in the agent himself, that is, in his essential self or identity. This is related to what I said earlier about unity – that is, that the agent *himself*, not just part of him, is supposed to be what stands behind the action. Earlier philosophers, of course,

---

<sup>10</sup> Questions about the bearing of these two implications on non-human agency are complicated. For example, it is clearly inappropriate to hold non-human animals responsible for what they do, yet our responses to them are often personal in character. Is it justified or rational to like or dislike a non-human animal for the kinds of things that she does? I leave aside such hard questions here.

tried to ground both of these implications in a metaphysical story about the freedom of the will, sometimes understood in terms of the absence of any causation governing the decision process. Such stories seemed outrageous to philosophers with naturalistic inclinations. But they also seemed, most notably to David Hume, to pit the two implications against each other.<sup>11</sup> If the activity implication requires us to regard actions as uncaused, then *ipso facto* it requires us to regard actions as uncaused by features of the agent's identity, and so renders the identity implication mysterious. How does the action reflect anything about the agent if nothing about the agent causes it?

I've said that these two implications pose problems for a naturalistic account of agency. Of course philosophers who hold naturalistic theories do have some things to say about them, some of which I will come back to. But first I want to talk about the way the view that agency is normatively constituted handles these two implications.

### 1.3 Activity, Identity, and Self-Constitution

At first, the Kantian and Platonic conceptions of agency that I sketched earlier may actually seem like particularly unhelpful places to look for an

---

<sup>11</sup> See David Hume, *A Treatise of Human Nature* (2<sup>nd</sup> Edition edited by L. A. Selby-Bigge and P. H. Nidditch; Oxford: The Clarendon Press, 1978) Book II, Part 3, Section 2, pp. 411-12; *An Enquiry Concerning Human Understanding* (3<sup>rd</sup> edition edited by L. A. Selby-Bigge and P. H. Nidditch, Oxford: The Clarendon Press, 1975) pp. 98-99.

explanation of the activity implication. Kant's view suggests that you are spontaneously active when you act in accordance with a law you make for yourself, but "making a law for yourself" sounds like it already describes an action. Plato's attempt to model individual action on collective action seems to face a similar problem. When you read the *Republic*, the parts of the Platonic soul seem like homuncular agents, who do various things, stand in various relations, and even exhibit various virtues and vices of their own. So these philosophers may seem to be guilty of attempting to explain the activity implication by positing little active agents as parts within the agent as a whole.

The story I tell in *Self-Constitution*, according to which we *constitute* our own agency or activity by adopting certain laws to govern our actions, is intended to be responsive to this difficulty. To see how it is supposed to work, start from a claim that I made in the previous section: the activity implication, although it seems like a metaphysical claim, also seems to be associated with a normative fact – the fact that agents can succeed or fail. I called that a normative fact, because to suggest that agents can succeed or fail seems to imply that they are under some kind of norm or principle which they may live up to or not. You might at first think that by itself this has no tendency to suggest that agency is normatively *constituted* – that is, to show that agency *consists* in following certain norms or principles. After all, agents are not the only things that can succeed or fail – your lungs, your kidneys, your heart, your car engine, and your alarm

clock can succeed or fail too, and they do not follow norms or principles. We regard them as capable of success and failure insofar as we regard them as functional objects, objects with some sort of job to do. To this extent, we do think of them as being under some sort of norm: that is, we apply some sort of normative standard to them. The standard in these cases can even be said to be a constitutive one, since a heart that does not pump blood fails *as* a heart, and an alarm clock that doesn't wake you up at the time for which you set it fails *as* an alarm clock. But this does not tempt us to think that the efficacy of such objects is constituted by their following these norms. Their efficacy is constituted, rather, by certain mechanical processes that enable them to do what the norm says they should do. Nor of course does it tempt us to believe they are agents in the sense we are looking for here – spontaneously active, originary beings.

But what makes *an agent* susceptible to success and failure in the sense we want here is not just the fact that we assign some purpose to his movements, and therefore hold him liable to a normative standard, as it were, from the outside.<sup>12</sup> Agency is liable to a standard of success and failure from the inside, that is, by virtue of its own nature: such liability is part of its constitution. An agent does not just effect something, the way the heart pumps blood or the engine keeps the car going – an agent *undertakes* to effect something, and to undertake to effect

---

<sup>12</sup> Here I ignore certain complications about how these reflections might apply to the efficacy of non-human animals. I discuss animal agency in *Self-Constitution*, Chapter 5.

something is to set yourself up for success or failure. Obviously, part of what it means for the agent to succeed is that the end the agent undertakes to effect does indeed get effected, and gets effected as a result of the agent's own movements. But the idea of efficacy that we associate with agency is not exhausted by the idea of actually bringing about the end, since an agent who brought about his end only accidentally – say by a deviant causal pathway – would still have failed as an agent. If I fire my gun wildly astray, but the bullet ricochets off a cast iron fence and happens to hit the target in exactly the spot that I intended, you will not cry “good shot!” To be successful in action is not merely to do something that brings about your end. To be successful in action, you must be *the kind of thing* that achieves the kind of end you are trying to achieve. Of course, an alarm clock that rings at the time for which you set it accidentally – say because it pops a spring at that very moment – is not successful as an alarm clock either. To be successful, the clock too must be the kind of thing that rings at the right moment. But agency requires even more than that. To act is not just to cause an end, but to *make yourself into* the cause of the end, and so to *make yourself into* the kind of thing that achieves that end. To be an agent is to transform yourself into a certain kind of cause. The activity we exhibit in action is a kind of *self-determined* efficacy.

The problem I started out from was that “making a law for yourself” seems to be something we do, that is, to be an exercise of activity, and that therefore Kant cannot appeal to that idea in order to explain the sense in which

we are active. What I have added now is an account of why activity amounts to a form of self-determination. How does this help with the problem?

It helps because of the connection between the idea of making a law for yourself and the very idea of being self-determined. To make yourself into a certain kind of thing just is to take a certain law as the law governing your actions. According to Kant, to be self-determined is precisely to be autonomous, that is, to engage in conduct that is governed by laws that you give to yourself, rather than letting the laws of nature determine what kind of thing you are. Kant argues that the law of autonomy is the categorical imperative itself. The categorical imperative in effect says: act in accordance with a law you impose on your own conduct; be self-determined. Since the laws we impose on our own conduct are supposed to be laws of action – that is, laws that determine the ways in which we are efficacious in the world – the laws that we impose on our own conduct must also conform to the hypothetical imperative: they must be laws directing us to take effective means to our ends. So putting those two standards together, the Kantian imperatives tell us that we must make ourselves into the kind of thing that tends to achieve our end. So the idea of agential activity – of self-determined efficacy – is captured by the content of these two Kantian norms. That means that the way we constitute our agency is by following those norms.

Let me try to put this a little more clearly; or maybe a little less. According to Kant, you have to conceive of yourself as an agent. That's a

conception of yourself that inevitably belongs to the deliberative standpoint. Since you have to conceive of yourself as an agent, you have to conceive yourself to be under the categorical and hypothetical imperatives, the norms of autonomy and efficacy. Now suppose – of course this is controversial – but suppose that we could work out exactly *what* a self-determined being would do. That is, suppose we could work out *which* laws a self-determined being would follow, as Kant supposes we can. That is the view that the categorical imperative yields determinate content. Suppose also that you can be motivated to follow those laws yourself, not incidentally but precisely because your conception of yourself as an agent is a conception of yourself as bound by those laws. Then you would be effectively self-determined. To put it rather bluntly, if you can be motivated by the thought that you are an agent and so bound by the laws of agency to act in exactly the way a genuinely active being would, then you are, effectively, active. No doubt there is some causal story about how you arrived at these thoughts, but that does not matter.<sup>13</sup> It is the *content* of the principles we follow, not the metaphysical story about how we came to follow them, that explains the activity implication. It does not matter how it comes about that you do it: by following the categorical imperative you render yourself self-determined and by following

---

<sup>13</sup> In *Self-Constitution*, §7.5.3 (pp. 154-58), I discuss Kant's view that the people in a republic can be genuinely sovereign regardless of how they came to be that way, so long as they have the correct forms of government. I take the point to be analogous: freedom is a matter of the content of one's form of government, not of its causal origins.

the hypothetical imperative you render yourself efficacious; taken together, by following these imperatives you render yourself an active being, one whose efficacy is self-determined. To put it yet another way, the categorical and hypothetical imperatives taken together give us the *form* of genuine activity, in the sense of the form of an exercise of the agent's self-determined efficacy. By conforming to these norms, we impose that form on our own actions.

The identity implication then follows, and in this story, unlike in the traditional free will story, the two implications are not at odds. As I said a moment ago, the categorical and hypothetical imperatives are formal principles in the sense that they give our actions the *form* of self-determined efficacy, which is the form of agency. But the principles in accordance with which we actually choose – our maxims, in Kant's language – also have some particular content or material.<sup>14</sup> As I said above, when we undertake an action, we make ourselves into *the kind of thing* that brings a certain end about. The content or material of our principles determines *what* kind of thing we make ourselves into, what kind of causes we are. In the course of choosing our actions, we constitute our own practical identities, as I have elsewhere called them. You make yourself into an

---

<sup>14</sup> Above I said it was the content of our principles, not their origin, that explained the activity implication. This might give rise to confusion. In those remarks, I meant the content of the two formal principles themselves – their content is that they impose the form of self-determined or autonomous efficacy on our actions. Now I mean the specific content of the principles on which those forms are imposed.

effective friend, teacher, parent, citizen, or whatever, by imposing the form of activity on principles derived from those roles. So the selves whom we are rendering active, if I may put it that way, are the possessors of those very practical identities. That is why our actions are the appropriate grounds of the kinds of normative and personal attitudes that are supposed to be responsive to a person's identity.<sup>15</sup>

So the idea is that by choosing our actions in accordance with certain norms, we render ourselves the kind of active beings whose movements can be said to have their sources in the self, beings to whom it is appropriate to respond with such reactions as love and hate, liking and disliking, gratitude and resentment. If these arguments work, the activity and identity implications are supported by the idea that action is normatively constituted – that is to say, by the idea of self-constitution. We make ourselves into agents by following norms that express the essence of self-determined efficacy, and we make ourselves into

---

<sup>15</sup> While I claim that this conception of action supports the identity implication, I think it should be clear that it doesn't support as "deep" a notion of responsibility as the traditional free will story aspired to do – nothing that would justify, say, eternal damnation. For that, it does matter how (and whether) it comes about that we think thoughts about self-determination: whether we have a reasonable moral education, say, and learn to think of ourselves as beings who can take charge of our own identities. To my mind, this is all to the good. We do not all have the same opportunities for self-constitution; responsibility is not that deep.

the particular agents who we are by the content that we give to those formal norms.<sup>16</sup>

#### 1.4 Activity and Identity in Naturalistic Accounts

Now let's return to the question how a naturalistic account of action might explain the activity and identity implications.

The naturalistic account must explain the activity implication by arguing that we exhibit our active nature when we act on the right kind of mental state: that is to say, it must be something about the particular mental state that causes action that explains the activity implication. We can call this mental state, whatever exactly it might be, an "an intention." But of course the idea of "forming an intention" gives rise to the same apparent difficulty as Kant's idea of "making a law for yourself." Is forming an intention something you do, perhaps even something you do intentionally? Or, if it is something that happens to you, why should the efficacy of intention count as *your own* efficacy any more than that of any other mental state – say if you drop something because you are nervous? Of course, as I mentioned before, there is also a problem about the identity implication. Why exactly should the content of your intentions provide

---

<sup>16</sup> For some precursors to the arguments in this section, see Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford University Press, 2009), in general, but perhaps especially §§1.4.2-1.4.3, 4.3.2-4.3.3, 5.1. See also "The Myth of Egoism," Essay 2 in Korsgaard, *The Constitution of Agency* (Oxford: Oxford University Press, 2008), §1.3, pp. 82-4.

the proper basis for personal and moral responses like love, hate, gratitude, or resentment, that are supposed to be directed not just to your accidental features, but to your very identity or self?

One possible way of addressing these questions is suggested by the work of Michael Bratman. I am going to start by talking about the identity implication. In conversation, I once asked Bratman whether an “intention” which someone had acquired through post-hypnotic suggestion would be any less the person’s own intention because of that.<sup>17</sup> What I had in mind was that the way in which this intention was formed threatens the identity implication. After all, we would not ordinarily hold someone responsible for an action grounded in a post-hypnotic suggestion, or so we might at first think. But Bratman said no, this would be just as much the person’s own intention, and the person’s own action, as any other. One way of taking this answer is as showing that Bratman holds a rather literal-minded version of the what I have been calling the naturalistic view: an action is a movement that is caused or causally guided by the mind, and an agent is just the person whose mind does the causing. Since this intention, regardless of how it got into the person’s mind, *is* in the person’s mind, the resulting action counts as his own, and that is all there is to say.

But in certain places Bratman says things that suggest a different possible account of the identity implication, one that appeals to the role of intentions in

---

<sup>17</sup> This was when Bratman visited my seminar on the Philosophy of Action in the Spring of 2009.

establishing a Lockean (or Parfitian) form of personal identity over time.<sup>18</sup> Bratman emphasizes the way that interlocking and long-term intentions do play a role in establishing our transtemporal identities. One can argue that the reason they play that role is because of the norms to which intentions commit us – norms of taking the means to your ends and of consistency of principle. Your intentions count as representative of yourself, of your identity, on the account I have in mind, precisely because continuing and interlocking intentions are such an essential part of what unifies you into a single coherent agent over time – that is, into a self. In other words, they play an essential role in self-constitution.

And on reflection we can see that my objection to supposing that an intention installed by post-hypnotic has its source in the agent's self or identity is not after all wholly based on its origin. We are also naturally inclined to envision the effects of a post-hypnotic suggestion as ephemeral, and we might think differently about it if we expected it to have long-term consequences for the person's identity. Suppose the post-hypnotic suggestion is not, for example, "when you wake up, kiss the first person you meet," but, say, "when you wake up, fall in love with the first person you meet, and never change your mind." Then, like Demetrius in *A Midsummer Night's Dream*, the person is changed

---

<sup>18</sup> See for instance his "Reflection, Planning, and Temporally Extended Agency" and, for a discussion of these issues that brings in my own work, "Two Problems about Human Agency," both in *Structures of Agency* (Oxford: Oxford University Press, 2007).

forever by the outside intervention.<sup>19</sup> I think we are rightly puzzled about how to think about these cases, which is why the situation of Demetrius at the end of the play leaves us feeling a little anxious. (A less pretty real-life example is the person whose character is changed for the worse, and permanently, by a brain injury.) In any case, if we read Bratman this way, we could read him as saying that *by* conforming to the norms consequent upon intention, we constitute our identities, and so we make ourselves into their authors, just as I want to say that by the way we choose our maxims – in accordance with the Kantian imperatives – we make the actions expressed in those maxims our own. We turn ourselves into the selves who then count as their authors. The naturalistic view, then, can explain the identity implication by adopting the self-constitutionalist view that our identities are constituted by the choices that we make when those choices are governed by the laws of reason.

What can the naturalist then say about the activity implication? Some philosophers might take an explanation of the identity implication to settle the question of the activity implication as well. As I mentioned earlier, the two implications seem joined in the thought that when we act, our movements have their source in our self or identity. So one can explain the two implications together if one supposes that the idea that the agent is active is completely

---

<sup>19</sup> By William Shakespeare. Demetrius, who loves Hermia, is made to fall in love with Helena instead under the permanent influence of a magic potion.

explained by the idea that his movements have their source in an identity that he in a sense constructs for himself. I do not believe that by itself this adequately captures the element of self-determination needed for the activity implication, however, because the Lockean notion of self-constitution requires only that the agent conform to his principles, not that he chooses them, or stands in an active relation to them, himself. That is why in my own story the two implications have separate explanations: the activity implication is explained by the appeal to the fact that the two imperatives give our actions the form of self-determined efficacy, while the identity implication is explained by appeal to the idea that we constitute our own particular agency by imposing *that form* on the particular material maxims that make us who we are. But there is certainly room for the naturalist to incorporate a thought about self-determination into his account. On a naturalistic account, having an intention involves having the thought that you will *do* something, and that thought must have some content. If I am right, that content is that one will exhibit self-determined or autonomous efficacy with respect to one's movements and the ends they tend to bring about. But if a thought about self-determined or autonomous efficacy brings us, first-personally, under the Kantian imperatives, then our actions are normatively constituted after all.

The naturalist can explain the activity and identity implications, then, but at least from the first-person point of view, the most obvious way to do this is by

taking certain elements of the self-constitution view on board. This thought brings us to the role played by the first-person perspective in giving shape to our conception of agency.

## *Part II: Self-Consciousness, Reason, and Activity*

### 2.1 The Self and the Mind

According to the naturalistic conception of agency, we identify actions with movements that are causally guided by the *mind*. What I have called the activity and identity implications suggest a slightly different view, one that I have argued can be captured by the self-constitution view: that we identify actions with movements that have their sources in the *self*.

In Section 1.4 I tried to show some ways in which we might bring these two ideas together. Still, we might wonder where exactly that leaves the relation between the naturalistic and normative conceptions of agency. Are they really two conceptions of the same thing, a single phenomenon called “agency”? There are cases in which it appears to be our practice to employ certain normative concepts wherever we find that certain natural ones apply. Absent some special argument, for example, we treat everything that fits the concept of a living human being as if the concept of a moral and legal “person” (say, responsible being with rights) applies to it. Or, for another example, on the international scene we treat every group of people organized under some sort of coercive

power as if the concept “sovereign nation” applied to it. When we consider the attribution of agency third-personally, it may seem tempting to think we are operating in this way: simply employing the normative conception of agency whenever the natural one is in place.<sup>20</sup>

On the other hand, when we consider action from a first-personal point of view, it seems nearly impossible to tease the two conceptions apart. Indeed one way to bring the two conceptions together is to point out that from your own point of view, there is really no difference between having an intention and committing yourself (even if conditionally) to performing the intended act – in other words, between having an intention and making a law for yourself. On that showing, the intention itself is normatively constituted; from your own point of view, your having an intention *consists* in your having brought yourself under certain particular categorical and hypothetical imperatives, that is, certain laws. This fact suggests a deeper connection between the two views.

And of course, one obvious explanation of the fact that we identify movements that arise from the *mind* or from mental states with movements that have their source in the *self* appeals to the special relationship between the mind or mental states and what we call self-consciousness. We are aware of ourselves

---

<sup>20</sup> One might take it as evidence that this is what we are doing that we are inclined to take almost all action personally, even in cases where it makes little sense to do so. We may respond with rage, for instance, to a pesky horsefly or mosquito, but more likely with mere dismay to poison ivy, merely because the horsefly or mosquito appears to be doing something to us.

– that is, of our *selves* – as the subjects of our mental states in a way that we are not aware of ourselves as the subjects of circulation and digestion. That is why, someone might propose, we automatically identify our selves with our minds in a way that we do not automatically identify ourselves with our bodies and our bodily processes.

I think there is something in this, as I am about to explain, but it cannot be the whole answer. The point I have just made applies as much to our experiences as to our intentions – I am aware of myself as seeing the sky, say – but we do not identify with our experiences, or hold ourselves responsible for having them, in the same way that we identify with our intentions.<sup>21</sup> It also applies as much to ourselves as the subjects of our emotions and our desires as of our actions, but whether we identify with our emotions and desires or hold ourselves responsible for having them seems to depend on the case. So if this sort of argument is to show us why we identify our selves with the subject of our intentions, it must show us that the *way* in which we are the subjects of our intentions is different from the way in which we are the subjects of our experiences and perhaps of our desires and emotions – different in a way that justifies us in identifying the self-conscious subject who is the subject of our intentions with the active self.

---

<sup>21</sup> This remark is only true if we take “experience” in a certain way, as I will explain later.

But this shows us just how deep the problem is. When I raised the question about the activity and identity implications, I was in effect asking what makes it possible for us to lay claim to our actions, and attribute them to our active selves. The problem we are confronting now is that exactly the same issue arises about those mental states of which we are supposed to be the active subjects – the ones that cause action. We cannot just pick some of those states out and say “we are active when those are operative” – we have to explain why that should be so. The normative constitutionalist is committed to the following answer: in order for our mental states and attitudes to be expressions of our own activity, we must impose the forms of mental self-determination upon that activity, by following the norms of mental activity. Mental activity in general must be normatively constituted. I believe that this is true; in fact I believe that *the norms that constitute mental activity are the laws of reason*. That is what the laws of reason *essentially* are. So my claim is that mental activity is not just governed by rational norms – it *consists* in following rational norms. In the rest of this paper I try to explain why someone might hold this view.

## 2.2 Mental Activity and the Laws of Reason<sup>22</sup>

In order to explain what I have in mind, I will begin by distinguishing rational from non-rational agency. When a non-human animal acts, I believe, she is guided through her environment by means of a representation of that environment that incorporates both perceptual information and appropriate desiderative or aversive responses. What I mean is that, for the other animals, perceptual representation and desire or aversion are not strictly separate. The animal finds herself in a world that consists of things that are directly perceived *as* food or prey, *as* danger or predator, *as* potential mate, *as* child: that is to say, *as* things that are to-be-eaten, to-be-avoided, to-be-mated-with, to-be-cared-for, and so on. I like to describe this by saying that an animal's perception has *teleological content*: the objects she perceives are marked out as being "for" certain things or as calling for certain responses. I believe this because I think it is hard to see how perception could have been of any use to the relatively unintelligent animals in which it first evolved if something like this were not the case. Perception could

---

<sup>22</sup> Versions of the account that follows have appeared in *Self-Constitution*, especially Chapter 6; "The Activity of Reason" (*The Proceedings and Addresses of the American Philosophical Association*, Volume 83, No.2, November 2009; and forthcoming in *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by R. J. Wallace, Rahul Kumar, and Samuel Freeman. New York: Oxford University Press, 2011) and in "Reflections on the Evolution of Morality." (*The Amherst Lecture in Philosophy* 5 (2010): 1-29.

<http://www.amherstlecture.org/korsgaard2010/>)

not merely provide a primitive animal with theoretical information on the basis of which the animal has to figure out what to do, so it must be that it tells her what to do.

I am taking “rational action” to be the contrast here. The idea of rationality is often associated with the idea of self-consciousness, where self-consciousness is understood to be an awareness of oneself as a subject – as the subject of one’s own mental states. But many scientists believe that something like this form of self-consciousness exists in at least some of the other animals. One way in which this is supposed to be revealed is by the ethologist’s mirror test. In the mirror test, a scientist paints, say, a red spot on an animal’s body and then puts her in front of a mirror. Given certain experimental controls, if the animal eventually reaches for the spot and tries to rub it off, or looks away from the mirror towards that location on her body, we can take that as evidence that the animal recognizes herself in the mirror, and is curious about what has happened to her.<sup>23</sup>

It is a little difficult to articulate exactly why the mirror test is supposed to reveal an awareness of oneself as a subject. The animal grasps the relation between the image in the mirror and her own body. In so doing, she seems to show that she grasps the relationship between *herself* and her own body. For she grasps the relationship between two things, a certain physical body and – well,

---

<sup>23</sup> To date, apes, dolphins, elephants, and some birds have passed the mirror test.

what? – we can say “and herself” – but what exactly is the “herself” that she identifies with that body? Perhaps the idea is that what she identifies as *herself* is the self that is the subject of her experiences, for instance the one who sees the spot in the mirror, of whose existence she must then have some awareness.

Interestingly, however, even if this is right, and shows that the animal knows herself as the subject of her *experiences*, it does not yet show that the animal must be aware of herself as the subject of her *attitudes* – that is, say, of her beliefs, emotions, and desires. This suggests a possible division within this form of self-consciousness. An animal might be aware of her experiences and of herself as the subject of those experiences, and yet her attitudes might still be invisible to her, because they are a lens *through* which she sees the world, rather than being parts of the world that she sees.<sup>24</sup> In that case, she would still function in the way I have called “instinctive.” The experiences that she was aware of having would still be experiences of things as “to-be-eaten” “to-be-fled” “to-be-cared-for” and so on; and her responses to those things would still be governed by the teleological content of her experiences.

But our self-consciousness, human self-consciousness, does extend to our attitudes. We know of ourselves that we want certain things, fear certain things,

---

<sup>24</sup> It is easier to understand what I mean here when you are thinking about practical, evaluative attitudes. It sounds odd to think of beliefs as a lens through which we see the world. But they are, in the sense that an animal could be moved by one belief to take up another without having any awareness of making an inference.

love certain things, believe certain things, and so on. We are also aware of the potential influence of our attitudes on what we will decide to do. We are aware of the *potential grounds* of our actions, that is, of the ways in which our attitudes incline us to respond to the situations in which we find ourselves. A similar thing can be said about our beliefs – we are aware of the ways in which our perceptions influence our beliefs, and also of the ways in which some of our beliefs may influence us to form others.

There is a problem with the way I have just drawn the contrast I am interested in here. I have spoken as if there are two different kinds of mental states, experiences and attitudes, of each of which we are the subjects, but in slightly different ways. That is not quite correct, so now let me put the point in a more careful way. Almost every conscious mental state or attitude has two distinguishable components. One consists in the fact that the mind, or rather the animal, is aware standing in a relation to some object outside of itself. The other consists in the *way* the mind, or the animal, is relating to that object, what the mind is *doing* with respect to the object. After all, even *experiencing* an object is a way of relating to it, although it is one of the hardest questions of philosophy exactly what that way is.<sup>25</sup> So the difference I am proposing is really this: that a non-rational animal might be aware of herself as the subject of her experiences only in the sense that she is aware of standing in a relation to the object, but not in the

---

<sup>25</sup> You have to write, say, the *Critique of Pure Reason*, in order to give an answer to that question.

sense that she has any idea that relating and responding to the object is *doing* something, that it is the work of her own mind. All of that – the particular work her mind is doing in relating to the object in question – is, from the point of view of the non-rational animal, packaged into the perceived object itself. When we say that a spider is “creepy” for instance, we package our own response to the spider into the spider, as if the fact that the spider produces a certain shiver in us were an attribute of the spider. I am claiming that non-rational experience is in general like that. By contrast, we rational beings are aware, or can be aware, when we experience objects, that we ourselves are at work on them – that the activity of our own minds is giving shape to the character of our experiences. The reason for focusing on attitudes, as I did originally, is that it brings this distinction out clearly. If a terrified animal could talk, she might say, “*there is a frightening thing, a thing to run away from.*” Perhaps, if we do not build too much content into the idea of “seeing,” she might even say, “*I see a frightening thing.*” But a terrified human being says, or can say, “*I am afraid of that thing,*” or even, “*I want to run away from that thing.*” Those formulations voice our awareness that part of the way we are experiencing the world is coming from *us*, from the workings of our own minds. We are aware not just of things and of our relatedness to things, but of our mental activity with respect to them.<sup>26</sup>

---

<sup>26</sup> On my view, to say you have “a mental attitude” or that you are in a “mental state” is just a way of saying that you are representing the world to yourself in a certain way (either practically

This is not the place to speculate about how or why we developed this form of self-consciousness. But I do want to venture one idea about that. I would like to suggest that self-consciousness of this form arose when the teleological conception of the world lost its normative hold on us, or at least that these two developments are associated with each other.<sup>27</sup> For our awareness that our own mental activity is involved in the way we represent the world to ourselves is at the same time an awareness that those representations are open to question. The teleological content may still be there – the situation may still appear to us as one that is, say, to-be-run-away-from – but that appearance no longer inevitably governs us. That presents us with a problem: now we have to decide what to do, and in another kind of case, what to believe. We have to take charge of our own responses to the world, and so to become active with respect to them in a different way than the other animals are. In order to solve that problem that we need reasons, reasons to act and believe as we do.

Now there are two, or at least two, views about how we arrive at those reasons. On what I call a substantive realist view, the reasons are simply there; it is a matter of finding them, and finding out whether our natural tendencies to respond as we do match them or not. I am not going to argue against the realist

---

or theoretically or both) where “representing” is *doing* something. The states and attitudes are not really static entities, though our consciousness of them is a like a snapshot that makes them appear that way.

<sup>27</sup> I do speculate about this question in “Reflections on the Evolution of Morality.”

view on this occasion; I have done so elsewhere. Instead, I am going to describe an alternative, which has come to be called a constructivist view.<sup>28</sup>

On a constructivist view, there is another way to describe the problem that we face here: the problem is one of unity. The idea is that when we become aware that we are representing the world to ourselves, when we turn our attention away from what we perceive and onto the fact that what we are doing is perceiving, then there is a way in which our experience loses its unity. What was once simply given to us as a unified environment is now given to us as a kind of heap of perceptions or experiences, and it is now up to us to put them back together into a unified conception of the world. In a similar way where once upon a time we always knew what to do in response to a situation, our own possible responses are now given to us as a kind of heap of desires and fears and impulses, and it is up to us to put ourselves back together into unified agents.<sup>29</sup> The work of reconstituting our conception of the world into a unified conception, and of reconstituting ourselves into agents is, in my view, the work of reason. We can arrive at the principles of reason – at their content – by asking how we carry on this work. The principles of reason are the principles by means of which we unify, or re-unify, both ourselves and our conception of the world.

---

<sup>28</sup> These passages are lifted from “The Activity of Reason,” which also contains some of the arguments against realism mentioned in the text.

<sup>29</sup> See *Self-Constitution*, 6.2.1-6.2.5, pp.121, 6.4.1-6.4.2, pp.125-6.

I have already said something about why the idea of agency requires unity: an action is supposed to be a movement, or the effecting of a change, that is somehow backed by the agent as a whole, that comes from the agent as a whole. I hope it is clear enough for purposes of this paper how conformity to the hypothetical and categorical imperatives, taken together, enable you to achieve that unity. When you will an end, you determine yourself to bring about the end. Willing an end is different from just wanting it, since when you will an end, you commit yourself to staying on the track of that end in the face of possible changes in your circumstances, and resulting changes in your motivational state. If I will to exercise, for instance, as opposed to just wanting to, or hoping I will, I will to exercise even if I do not happen to want to. When I adopt a maxim of exercising as a law, when I determine myself to exercise, then instead of being a mere heap of desires and impulses, some of the them favoring exercise, many of them not, I make myself into *the kind of thing* that exercises. In that way, I achieve the kind of unity that is essential to agency.

The parallel point about theoretical reason comes in two steps. First of all, unity is needed for our conception of the world, because the business, the function, of a conception of the world is to enable you to find your way around in it and to act effectively in it. In order to conceive the world as the sort of place in which you can find your way around and act effectively, you have to conceive it as a unified place. What that means is that the *relations* between the various

things in the world can be traced and established. If we can say nothing about how two things or events or regions of space-time are related to each other, we cannot think of them as parts of a single unified world. If we cannot trace causal relations, in particular, we cannot act effectively, since we cannot take means to our ends. So it is the business of a conception of the world to establish these various relations. Further argument is required, of course, but I suppose that we may think of the relations in question as logical, spatiotemporal, and causal. Speaking very roughly, these are the relations established by the principles of logic and what Kant thought of as the *a priori* principles of the understanding.

But – and this is the second step – in unifying our conception of the world, we are also unifying our minds themselves, and unifying them in a way that makes us the *agents* of that conception – that is, in a way that makes us active knowers. This is because the unity of the mind and the unity of its object are interdependent. Unless we conform our beliefs to logical and rational principles, our minds themselves are a mere heap of unrelated ideas that cannot really qualify as beliefs. A mere heap of unrelated ideas is not *about* anything, and therefore cannot count itself as thinking *about* anything or knowing anything. So our conception of ourselves as thinkers, and as possible knowers of a world independent of our minds – our conception of ourselves, if you will, as mental agents – depends on our idea that we can conceptualize the world as a unified object. As I have just suggested, our ability to do that depends on the

deployment of rational principles. Our unity as believers or knowers, or conceivers of the world, therefore, is also normatively constituted. To be an active knower is to follow norms that unify you into the active subject of a unified object – a knowable world.

If all of this is correct, belief and intention are not states that, as it happens, we can bring under the government of reason because they are active states. Rather, belief and intention are active states because they are essentially, constitutively, under the government of reason, and the laws of reason are the laws of mental activity. Mental activity itself is normatively constituted. To be active, and to be under the norms of reason, are one and the same thing.

### 2.3. Conclusion: Natural, Normative, and Collective Agency

Now let me return, finally, to the two questions I raised at the beginning of this paper – the question about the relationship between the natural and the normative conceptions of agency, and the implications of that relationship for questions about collective agency. I have claimed that a certain form of self-consciousness, a certain way in which we are conscious of ourselves as the subjects of our own mental attitudes, carries with it the knowledge that our own mental activity makes a contribution to the way the world appears to us, both theoretically and practically. This knowledge makes it both possible and necessary for us to take control both of our decisions and of the way in which,

ultimately, we represent the world to ourselves. We are faced with the task of forming a unified conception of the world, and at the same time, constituting our own mental unity as the active knowers of that conception. We are faced with the task of constructing a unified self that can stand behind the movements that it causes and claim them as the effects of its own activity, and so at the same time, constituting our agency. The way that we carry out these tasks is by following rational principles, the principles of theoretical and practical reason respectively.

But for whom, exactly, does this problem arise? One obvious answer is that it arises for a natural individual, an individual human being who is self-conscious in the sense I have been discussing. That is, it arises for a being who has a unified mind in the natural sense – the sense in which each of us has his own mind – but who is also conscious of his own mental activity. I am certainly not going to advance a view about what exactly unifies the mind of a natural individual. I do not think it is obvious; that is a whole topic in its own right. One obvious requirement is memory; another is that the faculties of sensation, perception, and representation more generally operate co-consciously, or in some way that brings their results into a common sphere. My point here is that a self-conscious natural individual with a unified mind in *that* sense, will inevitably be faced with the task of normatively unifying his agency and his conception of the world.

But the unity of agency is not *given* with the unity of the mind of the natural individual. Rather, what is given with the mind of the natural individual is the *need* for constructing the unity of agency and of one's conception of the world. So the relation between the natural conception of agency and the conception of agency as normatively constituted is this: being a self-conscious individual with a unified mind in the natural sense – being the seat of a unified consciousness – sets you the task of constructing a unified mind, and a unified agency – in the normative sense.

And – finally coming to the relevance of all this to the issue of genuinely unified collective agency – there is no reason to suppose that the possession of a unified mind in the natural sense is the *only* thing that sets us the task of normative unification. We can think alone, but we can also think together: we can talk. We can form our conception of the world alone, but we can also do that together. We can act alone, but we can also act together. In each case, we constitute ourselves the joint agents of the enterprise by observing the rules of normative unification together. On the view of action as normatively constituted, unification is an achievement, not a given. And means there is nothing that prevents people from acting together in the most literal sense of the word.

*Bibliography*

Bratman, Michael. *Faces of Intention: Selected Essay on Intention and Agency*.

Cambridge: Cambridge University Press, 1999.

Bratman, Michael. *Intentions, Plans, and Practical Reason*. Cambridge, MA:

Harvard University Press, 1987; rpt. CSLI Publications, 1999.

Bratman, Michael. *Structures of Agency: Essays*. Oxford: Oxford University Press,

2007.

Davidson, Donald. *Essays on Actions and Events*. Oxford: Oxford University

Press, 1980.

Gilbert, Margaret "Shared Intention and Personal Intention," *Philosophical*

*Studies* 144: 167-87, 2009.

Gilbert, Margaret. "What is it for Us to Intend?" in *Sociality and Responsibility*.

Lanham, MD: Rowman & Littlefield, 2000.

Hume, David. *Enquiry Concerning the Principles of Morals*, in *David Hume:*

*Enquiries Concerning Human Understanding and Concerning the Principles of*

*Morals*. Third Edition edited by L.A. Selby-Bigge and revised by P.H.

Nidditch. Oxford: Oxford University Press, 1975.

Hume, David. *A Treatise of Human Nature*. Second edition edited by L. A. Selby-

Bigge and revised by P. H. Nidditch. Oxford: Clarendon Press, 1978.

Kant, Immanuel. *Critique of Practical Reason* (Cambridge Texts in the History of Philosophy). Translated and edited by Mary Gregor with an Introduction by Andrews Reath. Cambridge: Cambridge University Press, 1997.

Kant, Immanuel. *Critique of Pure Reason*. Translated by Norman Kemp Smith. New York: Macmillan, St. Martin's Press, 1965.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals* (Cambridge Texts in the History of Philosophy). Translated and edited by Mary Gregor with an Introduction by Christine M. Korsgaard. Cambridge: Cambridge University Press, 1998.

Kant, Immanuel. *The Metaphysics of Morals* (Cambridge Texts in the History of Philosophy). Translated and edited by Mary Gregor with an Introduction by Roger J. Sullivan. Cambridge: Cambridge University Press, 1996.

Korsgaard, Christine M. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.

Korsgaard, Christine M. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009.

Korsgaard, Christine M. "The Myth of Egoism" in *The Constitution of Agency*. Oxford: Oxford University Press, 2008.

Korsgaard, Christine M. "The Activity of Reason." *The Proceedings and Addresses of the American Philosophical Association*, Volume 83, No.2, November 2009; and forthcoming in *Reasons and Recognition: Essays on the Philosophy of T.*

*M. Scanlon*, edited by R. J. Wallace, Rahul Kumar, and Samuel Freeman.  
New York: Oxford University Press, 2011.

Korsgaard, Christine M. "Reflections on the Evolution of Morality." *The Amherst Lecture in Philosophy* 5 (2010): 1-29.

<http://www.amherstlecture.org/korsgaard2010/>

Plato. *Republic*. Translated by G. M. A. Grube, revised by C. D. C. Reeve, in  
*Plato: Complete Works*. Edited by John M. Cooper. Indianapolis: Hackett  
Publishing Company, 1997.

Rousseau, Jean-Jacques. *On the Social Contract*. In *The Basic Political Writings of Jean-Jacques Rousseau*. Translated by D.A. Cress. Indianapolis: Hackett  
Publishing Company, 1987.

Shakespeare, *A Midsummer Night's Dream*. in *The Riverside Shakespeare*, 2<sup>nd</sup>  
edition edited by G. Blakemore Evans and J. J. M. Tobin. Boston:  
Houghton Mifflin, 1996.

Velleman, J. David. *The Possibility of Practical Reason*. Oxford: Oxford University  
Press, 2000.

Velleman, J. David. *Practical Reflection*. Princeton, Princeton University Press,  
1989.