



# Philosophical Explorations

An International Journal for the Philosophy of Mind and Action



ISSN: 1386-9795 (Print) 1741-5918 (Online) Journal homepage: <https://www.tandfonline.com/loi/rpex20>

## Constitutivism and the virtues

Christine M. Korsgaard

To cite this article: Christine M. Korsgaard (2019) Constitutivism and the virtues, *Philosophical Explorations*, 22:2, 98-116, DOI: [10.1080/13869795.2019.1599050](https://doi.org/10.1080/13869795.2019.1599050)

To link to this article: <https://doi.org/10.1080/13869795.2019.1599050>



Published online: 05 Apr 2019.



Submit your article to this journal [↗](#)



Article views: 651



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 7 View citing articles [↗](#)

## Constitutivism and the virtues

Christine M. Korsgaard\*

*Department of Philosophy, Harvard University, Cambridge, US*

In *Self-Constitution*, I argue that the principles governing action are “constitutive standards” of agency, standards that arise from the nature of agency itself. To be an agent is to be autonomously efficacious, and the categorical and hypothetical imperatives arise from those two attributes. These principles are also “constitutive” of agency in two more specific ways. First, they meet the “constitution requirement”: the object must meet the standard in question, at least to some extent, in order to be the kind of object that it is. Second, they meet the “self-constitution requirement”: the object *makes itself into* the kind of object that it is *by* conforming to the standard. That is, the agent makes herself into an agent, and into the particular agent who she is, by conforming to those standards. Some neo-Aristotelians believe that Aristotelian virtues are constitutive standards. In this paper, I first ask why moral philosophers should focus on the virtues at all, considering the views of David Hume, Philippa Foot, and Aristotle. I then ask whether Aristotelian virtues meet the constitution requirement, and suggest that there are grounds for this view in the *Nicomachean Ethics*. But Aristotelian virtues do not meet the self-constitution requirement, which leaves Aristotle unable to explain moral responsibility. I end by examining the role that Aristotelian virtues could play in a Kantian ethic.

**Keywords:** action; agency; Aristotle; constitutive standard; Philippa Foot; David Hume; Immanuel Kant; responsibility; self-constitution; virtue

### 1. Introduction

1.1 A familiar view holds that some of the norms under which an object falls are given by the object’s nature, and may be justified by that fact. We can call these “internal standards.” Cheetahs should be swift, tigers should be ferocious, and antelope should be cautious. Knives should be sharp, encyclopedias should be accurate, and cars should be easy to handle – that is to say, they should be responsive to the manipulation of the steering wheel. A little less obviously, perhaps, hotel rooms should be clean, airports efficient, and grocery stores should offer a variety of brands. To make judgments of this kind, we need only have a functional or teleological conception of the nature of the object in question. That is given, in my first three cases, by an Aristotelian or Darwinian conception of organisms according to which their function is to maintain themselves and their species by staying alive and reproducing. In the other cases, it is given by a more or less instrumental view of what the objects are for, or of the role that they play in our lives.

---

\*Email: [christine\\_korsgaard@harvard.edu](mailto:christine_korsgaard@harvard.edu)

In *Self-Constitution*, I argue that the principles of practical rationality are what I call “constitutive” principles of action, or more precisely of agency. (Korsgaard 2009a) To be an agent is to be the autonomous and efficacious cause of events in the world. In order to be efficacious, you must conform to the hypothetical imperative, and in order to be autonomous, you must conform to the categorical imperative. By conforming to these principles, I argue, you make yourself into an agent: you *constitute* yourself as an agent.

Norms that are “constitutive” in my sense are internal standards, but the idea goes beyond that in two ways. The first is that unless the object conforms to the standard, at least to some extent, it ceases to be the kind of object that it is. I’ll call that the “constitution requirement.” The second is that the object *makes itself* into the kind of object that it is *by* conforming to the standard. I’ll call that the “self-constitution requirement.”

In this paper I discuss the relation between constitutivism and the virtues. In Part 2, I discuss the relation between the constitution requirement and the idea of a virtue. I first raise a problem about why moral philosophers should focus on the virtues at all. I use David Hume’s version of virtue ethics to show that there is a sense in which talk of the virtues is otiose, and can be misleading. Hume nevertheless does have a good reason for focusing on the virtues – he thinks that our approval of a person’s motivational dispositions is what *makes* them virtues, what confers a moral value on them. Philippa Foot’s theory of natural goodness shares the problem without sharing Hume’s reason for incurring it. I argue that there is a reason, different from Hume’s, why Aristotle himself focuses on the virtues, namely, that there is a genuine sense in which in Aristotle’s theory an agent’s virtues and vices constitute his will. But this feature of Aristotle’s theory depends on a feature of his moral psychology that Foot does not clearly endorse.

With that in view, in Part 3, I turn to a discussion of the importance of the self-constitution requirement for giving a proper account of the nature of action, in particular of why actions are the sorts of things for which we hold people responsible. I argue that Aristotle’s view does not meet that requirement, while Kant’s view does. Finally, in Part 4, I consider the role that the virtues, on an Aristotelian conception, might play in a Kantian ethical theory.

## 2. Virtue and the constitution requirement

2.1 According to the constitution requirement, conforming to the standards that apply to an object, at least to some extent, is what *constitutes* the object as the kind of object that it is. This is not to say that things cannot violate the constitutive standards that apply to them, and still be the kinds of things they are. But there is a limit – if the object falls too far below the standard, the object ceases to exemplify the kind. A standard is constitutive in this specific sense not merely if it arises in some general way from the nature or function of the object, but if it arises from (something like) the very concept of the object. A house is, by definition, a kind of shelter. It may have a leaky roof and still be a house, but a thing that provides no protection from the elements – a thing that has altogether ceased to shelter – is no longer a house.<sup>1</sup>

Some of the examples of standards that I gave in the opening section of this essay do not, or do not obviously, conform to the constitution requirement. It certainly seems that there could be a grocery store that offers only one brand, although we might find it disappointing, and not only that, but disappointing *as* a grocery store. “Selection” is one of our criteria for good stores these days. There’s a similar point to make about the other two members of my last group of cases. When I first appealed to the idea of a constitutive or internal standard, in my attempts to explain the normativity of the hypothetical imperative, I contrasted two possible normative standards one might apply to cakes: tasting good and being ten feet tall. (See Korsgaard 2008a, 61) The first, tasting good, I claimed, is internal or constitutive: a thing that

does not taste good at all just is not a cake, or in any case, if you are not trying to make it taste good, you are not really making a cake.<sup>2</sup> The second, being ten feet tall, by contrast, is an external standard: it involves an extraneous purpose one might happen to have for a cake. For (an admittedly dated) example, you might want it to be large enough to accommodate a showgirl, who will pop out and sing happy birthday when the right moment comes. But the cleanliness of a hotel room or the efficiency of an airport does not seem to fit either of those categories: the value of cleanliness does not spring from an extraneous purpose one happens to have for an hotel room, nor does the value of efficiency spring from an extraneous purpose one might have for an airport. Yet these standards seem to fall short of springing from the concept of a hotel room or an airport. The difficulty is not just that dirty hotel rooms and inefficient airports can exist. It is true in general, or so I claim, that being subject to a constitutive standard is compatible with failing to meet the relevant standard *to some extent*: as we saw, a house may have a leaky roof. But in that case, as I said before, there is a limit below which the object cannot fall without ceasing to be the kind of object that it is. But it is hard to apply the idea of such a limit in the cases I am talking about now. We have some idea what it means for an object to fall so far short of sheltering that it no longer counts as a house, but I doubt whether we can assign any clear meaning to something's being so dirty that it no longer counts as a hotel room. So a standard may be internal without being constitutive.

2.2 "Clean" and "efficient," in this context, are virtue terms, meaning, according to Plato and Aristotle, that they signify that the object has properties that enable it to perform its function (*ergon*) well.<sup>3</sup> Many such terms have a gestural quality. What I mean is that they gesture at the fact that the object has other properties, sometimes more naturalistically describable, that more directly constitute its well-functioning. I gave an example of that in my opening paragraph, when I said that a car's "handling well" means that it is responsive to the manipulation of the steering wheel. To say that the car handles well is to gesture at the fact that it has naturalistic functional properties like that. In my other examples, I named the naturalistic functional properties more directly: "sharp," "swift," "cautious." In calling these terms "naturalistic" I don't mean to deny that in context, they too are virtue terms. Indeed, I don't think there's a clear divide here, only that different virtue terms stand in more or less direct relation to what really matters to an object's being a good object of its kind, which is that it should function well. That is, as it were, the ultimate virtue of an object.

2.3 There is an important case in which the use of virtue terms is gestural, namely, the case in which we use them to refer indirectly to properties of the will, that is, to an agent's principles.<sup>4</sup> In fact this use can be a source of philosophical confusion, as I will now show.

Consider Hume's famous argument that what makes an action morally good cannot be the fact that it is done from the motive of duty. Hume says:

'Tis evident, that when we praise any actions, we regard only the motives that produced them, and consider the actions as signs or indications of certain principles in the mind or temper. The external performance has no merit. We must look within to find the moral quality ...

After the same manner, when we require any action, or blame a person for not performing it, we always suppose, that one in that situation shou'd be influenc'd by the proper motive of that action, and we esteem it vicious in him to be regardless of it. If we find, upon enquiry, that the virtuous motive was still powerful over his breast, tho' checked in its operation by some circumstances unknown to us, we retract our blame, and have the same esteem for him, as if he had actually perform'd the action, which we require of him.

It appears, therefore, that all virtuous actions derive their merit only from virtuous motives, and are considered merely as signs of those motives. From this principle I conclude, that the first virtuous motive, which bestows a merit on any action, can never be a regard to the virtue of

that action, but must be some other natural motive or principle ... Before we can have such a regard, the action must really be virtuous; and this virtue must be derived from some virtuous motive: And consequently the virtuous motive must be different from the regard to the virtue of the action.<sup>5</sup>

Hume concludes that a desire to do what is virtuous by itself cannot give us any guidance. Before we can do what is virtuous because it is virtuous, we must deem certain actions virtuous, and that must be because they characteristically spring from motives that we deem virtuous – motives other than the desire to do what is virtuous. To put the point in Kantian terms, Hume thinks that the motive of duty by itself has no content.

Philippa Foot closely echoes these thoughts in *Natural Goodness*:

What, for instance, distinguishes a just person from one who is unjust? The fact that he keeps his contracts? That cannot be right (*Hume*: “the external performance has no merit”), because circumstances may make it impossible for him to do so. (compare *Hume*: “If we find, upon enquiry, that the virtuous motive was still powerful over his breast, tho’ checked in its operation by some circumstances unknown to us, we retract our blame ...”). ... ‘Of course,’ someone will say at this point, ‘it is the just person’s intention, not what he actually brings about, that counts.’ (compare *Hume*: “we must look within to find the moral quality.”) But why not say, then, that it is the distinguishing characteristic of the just that for them certain considerations count as reasons for action, and as reasons of a given weight? Will it not be the same with other virtues, as for instance the virtues of charity, courage, and temperance? Those who possess these virtues possess them insofar as they recognize certain considerations (such as the fact of a promise, or a neighbour’s need) as powerful, and compelling, reasons for acting. (compare *Hume*: “we always suppose, that one in that situation shou’d be influenc’d by the proper motive of that action”).<sup>6</sup>

I’ll come back to Foot. First I want to explain where Hume goes wrong.

Hume contrasts placing the value of an action in its “motive” or “within” with placing it in “the mere external performance.” Elsewhere in the *Treatise*, Hume identifies “motives” with things like “benevolence” or “malice” or “ambition,” qualities that he also deems to be virtues or vices. There are awkwardnesses on both sides of this inner/outer contrast.

First of all, no one thinks a *wholly* “external performance,” if that just means a bodily movement, has any moral value. Suppose that you are starving, and I am about to eat a sandwich when I learn about this. And suppose that just then I am attacked by a series of involuntary muscle spasms that cause me to make exactly the same physical movements I would make if I were giving you my sandwich. Of course no one would claim that this wholly “external performance” has any moral value. An act must be done with a certain proximate or immediate intention in order to count as an *act* at all. And that proximate or immediate intention is already part of the action’s motive. So in order to even count as “*giving* you my sandwich” I have to at least intend to transmit the sandwich from my possession to yours.

But of course one can still have different motives for carrying out a certain proximate or immediate intention: I might give you a sandwich, say, simply in order to alleviate your hunger, seeing that as worth doing for its own sake. Or I might give you a sandwich because I believe you will pay me for it and I hope to get the money, or because I dislike you and I am hoping you will choke on it. Or, as Hume thought of the matter, I might give you a sandwich because I think it is my duty to give you a sandwich. But this last case, at least according to Kantians, is where Hume gets it wrong. Here is why:

I just contrasted these cases:

- (1) I give you a sandwich simply to alleviate your hunger.
- (2) I give you a sandwich to get the money you will pay me for it.
- (3) I give you a sandwich hoping you will choke on it.
- (4) I give you a sandwich because it is I think it is my duty to do so.

There are two important points about this list of cases:

The first one has to do with the awkwardness on the “inner” side of Hume’s “inner/outer” contrast. Hume, as I mentioned, sometimes describes things like “benevolence” or “malice” or “ambition” as “motives.” But actually, there is a difference in the way we think about the idea of a “motive” when we think about it from the first person or the third person point of view. I have just said that any act has an immediate or proximate intention; when we think of motives for performing a certain act from the first-person point of view, it is natural to think of them as further or more fully specified intentions. So for instance when I say that I give you a sandwich in order to alleviate your hunger, or hoping you will pay me for it, or hoping you will choke on it, I am specifying the further ends I hope to achieve by transmitting the sandwich from my possession to yours. But of course I might have further intentions still: I might be hoping you will pay me for the sandwich because I hope to get rich, or because I need to pay a debt, or because I want to make a contribution to charity. We might reasonably suppose that from this first-person point of view, we have specified the agent’s motive only when we have reached what I will call his final intention, which will be his fully specified intention. We only get to the agent’s fully specified intention – his final intention – when we arrive at a description of the action under which the agent values it for its own sake. There is no further reason, let us say, why I want to alleviate your hunger; or to get rich myself – those are simply states of affairs I see as valuable. Everything else equal, therefore, they make it seem to me as giving you a sandwich is a thing worth doing.<sup>7</sup>

We think of things this way when we look at the action from the agent’s own point of view. We spell out his final intention in order to see what, *from his point of view*, makes the act worth performing. Once we do that, we have a somewhat different way of specifying “motives” from the third person. We describe someone who acts from a certain kind of final intention as being motivated in a certain way, or as having a certain kind of virtue or vice. That’s when we tend to use terms like “benevolent” or “ambitious.” So for instance, in my four cases, we might, as a first pass, specify the “motives” (or virtues and vices) this way:

- (1) I give you a sandwich simply to alleviate your hunger. (Benevolence)
- (2) I give you a sandwich to get the money you will pay me for it. (Self-interest)
- (3) I give you a sandwich hoping you will choke on it. (Malice)
- (4) I give you a sandwich because I think it is my duty to do so. (Duty)

On this view, strictly speaking, people do not act *from* things like benevolence, self-interest, or malice. Those ideas do not appear as part of the content of their deliberations and intentions. Rather, they act from their final intentions, and when they act from certain kinds of final intentions, we describe them with those words. The words are gestural, and what they gesture at is a type of intention. This, I take it, is part of Foot’s point when she insists that a just person is one who acts on certain kinds of reasons, such as say, the fact of a promise.

This brings us to the second important point about the list. Notice that “duty” is a little different from the other items, because, in the schema above, “duty” does appear in the

content of the intention: I give you a sandwich because I think it is my duty to do so. That suggests a contrast between someone who gives you a sandwich in order to alleviate your hunger and someone who gives you a sandwich in order to do his duty: a contrast between two different purposes one might have in giving someone a sandwich. But at least according to Immanuel Kant, that kind of contrast does not correctly capture the difference between someone who acts from duty and someone who does not. Kant thinks that the duty in question *is* to “give someone a sandwich simply in order to alleviate his hunger.” Satisfying the needs of others is what he calls an “obligatory end.” Your duty is not just to perform an act, but to act *with* a certain final intention.<sup>8</sup>

Kant’s view requires that there be a further element to an agent’s motivation, one that our schema so far has not captured. Kant thinks there is a difference between two agents, both of whom adopt the final intention of giving-you-a-sandwich-in-order-to-alleviate-your-hunger. One of them adopts this final intention simply because he *wants* to alleviate your hunger: that is the naturally sympathetic person in the famous example in the first section of the *Groundwork*.<sup>9</sup> The other adopts the final intention of giving-you-a-sandwich-in-order-to-alleviate-your-hunger because he supposes that it is his duty to do that act with that intention: that is, he sees the alleviating your hunger as an obligatory end, regardless of what he happens to want. Kant thinks these are two different ways of being motivated to adopt a certain final intention. So now we have the following contrast:

- (1) I adopt the final intention of giving-you-a-sandwich-in-order-to-alleviate-your-hunger because that is what I naturally want to do. (acting from natural sympathy or benevolent inclination)
- (2) I adopt the final intention of giving-you-a-sandwich-in-order-to-alleviate-your-hunger because I recognize that it is my duty to have this intention. (acting from duty)

It is this contrast that Kant works with in the argument of first section of the *Groundwork*. Kant argues that these two characters have the same purpose or final intention, but that they adopt the maxim of acting on that intention on different grounds. The first person is moved by the maxim’s content, while the second is moved by its universal form, that is, by the fact that it must be willed as a universal law.<sup>10</sup> On Kant’s view, the difference between the person who acts from natural benevolence or sympathy and the person who acts from duty does not rest in different final intentions, or in different purposes, as he puts it, but in the principle that governs the adoption of their final intentions.

Hume’s failure to distinguish first- and third- personal ways of describing motivation leads him to overlook the possibility that Kant exploits. For we have here arrived at a reply to Hume’s argument that the motive of duty cannot be the “first virtuous motive, which bestows a merit on any action.” At least if Kant is right to suppose that some maxims cannot be willed as universal laws, and therefore that their opposites must be so willed, then we *can* get content from idea of acting from duty. If there is this further aspect to our motivation, the grounds on which we adopt our final intentions, then Hume’s argument fails.

2.4 Hume thinks that benevolence is what he calls a “natural virtue”: the benevolent person is moved by thoughts about the other’s need, rather than thoughts about duty or virtue. Third-personally, we might call a person who tends to want good things for other people “benevolent.” But it is misleading to think of benevolence as an attribute that *explains* the tendency to want good things for other people. “Benevolence” is not a force at work in the person, but a third-personal name we give to people who tend to want

good things for other people for its own sake, without any further or ulterior intention. No doubt there is some explanation of the tendency to want good things for other people, some package of perceptual tendencies, saliences, a certain kind of empathy, and so on, and we can call that explanatory package “benevolence” if we like. But that again is gestural. What makes the person morally good, on Hume’s account, is the package of tendencies themselves, not their grouping under the heading “benevolence.”

So why should moral philosophers talk about “benevolence” at all? Why not just talk about which final intentions we should have? For Hume, talk about the virtues is not otiose because morality comes into the world through third-personal evaluative judgments. As a sentimentalist, Hume believes that it is our *approval* of that package of natural tendencies, whatever exactly they might be, that *gives* “benevolent” action its moral character.<sup>11</sup> On Hume’s account, our approval is what *makes* being motivated in that way morally good. We talk about motives in this third-personal way when we are talking about what is morally good because according to Hume, moral judgment itself is, at least in the first instance, a third-personal exercise; its role in first-person deliberation is secondary.

But if Kant is right, we should distinguish between moral and natural benevolence, and only the former is a virtue.<sup>12</sup> What makes a person benevolent in the moral sense is not the content of his desires, or the package of attributes that explains his having desires with that content, but rather the nature of his principles. So if we say that someone has the moral virtue of benevolence, we are using the virtue-term to gesture specifically at a property of his *will*, his principles. And of course according to Kant what makes his will good is not that we approve of him from a third-personal point of view. It is that his principles are the right ones, the ones that are picked out by the categorical imperative. On the constitutivist reading, that means that his principles are ones that enable him to function as an agent – that they meet the constitution requirement. In any case, in a Kantian account there is no special reason to talk about someone’s virtues, *if* that kind of talk only gestures at a person’s principles. The important thing is the principles themselves.

2.5 Now think about Philippa Foot. One might suppose that in granting that what makes a person “just” is the nature of the reasons on which he acts and the weight he assigns to them, she is granting that what makes an agent well-functioning is his principles. To say that someone is “just” is simply to say that he has adopted a certain kind of principle: in her words he has “accepted a certain group of considerations as reasons for action.” But then why does she want to talk about the virtues? If to say that someone has a virtue is just to give a third-personal name to the fact that he acts on certain principles, surely the principles are what moral philosophy should be concerned with. Foot is supposed to be an Aristotelian. She does not agree with Hume that what makes a property a virtue is that we third-personally approve of it; she is supposed to think that what makes a property a virtue is that it makes us fit for the human form of life. And yet she says, puzzlingly:

... the description ‘just’, as applied to a man or a woman, speaks of how it is with him or her in respect of the acceptance of a certain group of considerations as reasons for action. If justice is a virtue, this is what the virtue of justice rectifies, that is, makes good ... Similarly, if charity is a virtue, this is because it makes its possessor’s action good in the area of aims such as the relief of poverty. (Foot 2001, 12)

How can the virtuous nature of “justice” or charity “make it good” to act on certain reasons for action? On an Aristotelian view, what is supposed to make acting on these reasons good is the fact that acting on them fits us for the human form of life – makes us well-functioning

as human beings – not the fact that acting on them counts as virtuous. Either saying that a certain way of acting is virtuous adds something, or it does not. If it is just a way of recording the fact that someone has good principles, then we should talk about the principles, and what makes them the right ones to act on. The fact that acting on those principles counts as “virtuous” does no theoretical work.

2.6 In place of the Aristotelian notion of function, Foot puts the notion of the general life form of a species, as that would be described in what Michael Thompson calls “Aristotelian categoricals” – categorical statements of the form, “Vervet monkeys have three kinds of alarm calls,” or “Kangaroos live in groups of ten or more” or “Vultures are monogamous and mate for life.”<sup>13</sup> These generic claims, as has often been noted, are not universal or statistical claims: they are not falsified by the occasional promiscuous vulture or a group of kangaroos that has shrunk to eight. On Foot’s view, to have the virtues is simply to be well-suited to live in the way characteristic of your species, as that life form is described by these categoricals, or to have the properties that make you well-suited for that life.

Foot does acknowledge the role of function, as opposed to just the Aristotelian categorical, to this extent: she thinks the only Aristotelian categoricals that matter to judgments of natural goodness are the ones that have something to do with a species’ ability to function. Suppose a species characteristically has a certain pattern of coloring – Foot’s example is the blue spot on the head of the blue tit – but that that pattern plays no functional role in the bird’s life. (It does not, say, make the bird more likely to attract a mate.) A blue tit that lacked that spot would not be defective, since this difference would not get in the way of its leading a blue tit’s normal life. (See Foot 2001, 30–31)

Foot tells us that her account of human virtue makes judgments of human goodness continuous with the account of goodness and defect in plants and animals:

The structure of the derivation is the same whether we derive an evaluation of the roots of a particular tree or the action of a particular human being. The meaning of the words ‘good’ and ‘bad’ is not different when used of features of plants on the one hand and humans on the other, but is rather the same as applied, in judgments of natural goodness and defect, in the case of all living things. (Foot 2001, 47)

I think it is interesting that Foot focuses here on “features of plants” rather than the plants themselves, and compares those to the “particular actions of human beings.” When we say that a human being is virtuous or does virtuous things, we move easily to the conclusion that she is a good human being. But as Foot herself notes, we do not talk this way about organisms generally. (See Foot 2001, 66) We are not much inclined to say of an oak tree with “good roots” that it is therefore a good oak tree. The more natural thing to say is that the roots are good *for* the oak tree; we might say of the tree itself that it is healthy. In any case, if we did call the tree a good one because of its roots, there would be no real question about whether being a good oak tree is good for the oak tree. But there is a real question about whether being a good human being, in the moral sense, is good for the human being. Some contemporary Aristotelians seem to think the connection between being good and having a life that is good for you is obvious, but Foot herself takes Elizabeth Anscombe and Gavin Lawrence to task for that: “It is too quick to say that because human goodness belongs to those who have the virtues, human good is what they will attain in acting well.” (Foot 2001, 92)

In any case, Foot seems to think that once we have identified the Aristotelian categoricals that have something to do with good functioning, we can pretty much read the virtues off from them. For instance, to arrive at the conclusion that justice is a virtue, Foot gives us

the following Aristotelian categorical: “Humans establish rules of conduct and recognize rights.” (Foot 2001, 51) Since that is the way human beings live, and justice is a property that enables you to live that way, justice is a virtue. Her accounts of the other virtues, so far as I can see, would not turn out to be very different from the kinds of instrumental accounts that Hume might give: given that human beings have unruly appetites that must be controlled, temperance is a necessary property for living well in the human way, given that human beings have dangers to face, courage is a necessary property for living well in the human way, and so on.

2.7 That isn’t how it works in Aristotle. Aristotle’s view is that virtues make you good at the human function, or in Greek, *ergon*. Modern commentators on Aristotle sometimes urge that when Aristotle talks about a thing’s “*ergon*,” he means not just purpose, but also something like characteristic activity. If that means simply what human beings characteristically do, I think that Aristotle’s concept of an “*ergon*” is more specific than that. I think he means *how* a thing does what it does.<sup>14</sup> The three kinds of life Aristotle describes in his famous function argument represent three ways of going about the business of living: simply through vegetative processes, like plants, or also by means of perception and action, like animals, or more specifically by means of rational thought and action, which is what is distinctive of human beings.<sup>15</sup> A virtue in Aristotle’s sense must be a property that makes that specific way of functioning possible. Since what is characteristic of human functioning is that we are rational, the virtues that are specific to human beings must be properties that make us good at doing and believing things in characteristically rational ways. In the practical case, that means they must make one good at the deliberation leading to choice (*prohairesis*).<sup>16</sup> According to Aristotle, we cannot establish that any property is a *moral* virtue without establishing that one cannot *reason well* without this property.

Foot could say some of these things – she could say that the virtues make us good at practical reasoning – but she does not seem to mean them in the same way that Aristotle does. What she seems to mean is simply that unless you take the right reasons into account, you do not *count* as having the relevant virtue. But Aristotle thinks that acquiring the moral virtues is what makes us capable of reasoning well about practical matters. This is because of a specific feature of his moral psychology. Aristotle believes that our thoughts about the good are determined, or at least guided, by our pleasures and pains and the appetites and passions that go with them. This in turn is because he believes that pleasures and pains amount to a kind of *perception* of goodness and badness. He also thinks that these perceptions of goodness and badness move us directly. He says:

To perceive, then is like bare asserting or thinking; but when the object is pleasant or painful, the soul make a sort of affirmation or negation, and pursues or avoids the object. ... Both avoidance and appetite when actual are identical with this: the faculty of appetite and avoidance are not different, either from one another or from the faculty of sense-perception; but their being is different.

To the thinking soul images serve as if they were contents of perception (and when it asserts or denies them to be good or bad it avoids or pursues them.) That is why the soul never thinks without an image. (Aristotle 1984b, 3.7, 431a7–16.)

The moral virtues are states of the appetites and passions that, given that they play this perceptual role, contribute to our correct cognitive grasp of the good – to practical wisdom. This cognitive grasp of the good then in turn informs our perceptions of particular situations. Courage is necessary for practical wisdom because our ability to judge which risks are worth taking is subject to distortion by the amount of fear that we feel in the face of certain dangers; temperance is necessary for practical wisdom because our ability

to judge which of our appetites is worth satisfying is subject to distortion by those appetites themselves. It follows that all of our passions and emotions together inform a kind of normative conception of the world, in light of which we *perceive* certain situations as calling for certain normative responses and we are moved by those perceptions to act accordingly. This normative conception of the world and the normative perceptions to which it gives rise are the “moving principles” that makes our actions voluntary – that is, that makes them our own. (See Aristotle 1984a, 3.1.) We act *from* that conception, in the sense that it is what moves us.<sup>17</sup>

If that is right, then for Aristotle there is a sense in which a person’s virtues really *are* constitutive of her will. A person’s virtues and vices determine how she responds to situations by determining what she perceives as good and bad. That is why Aristotle focuses on the virtues. For him virtue terms are not just third-personal terms for gesturing at the fact that someone acts from the right reasons or principles. They designate properties or aspects of the will itself.

In fact, Aristotle takes these ideas a step further. Aristotle believed in the unity of the virtues, which in effect means that there is really only one virtue, or one virtuous state of the person, corresponding to the possession of practical wisdom.<sup>18</sup> Aristotle believed this because he recognized that no aspect of a thing’s well-functioning is really separable from any other. To affirm the unity of the virtues is to acknowledge that *any* of a person’s affective dispositions is *potentially* relevant to his judgment of what would be good to do in *any* situation. We typically think of someone as making the wrong choices in battle because of cowardice or rashness, but someone might make the wrong choice in battle because of greed, say by choosing to attack a place because it is worth looting rather than for its military importance. It is often noticed that if the virtues really are unified in this way, it is impossible to have one virtue without the others: no one is, say, courageous but intemperate. This means that if the virtues are unified, what a single virtue term refers to, strictly speaking, is the absence of a vice. The vices, unlike the virtues, are not unified, and can exist apart. So to say that someone is courageous is to say that if he makes the wrong judgment about what to do, it is not because of excess fear of certain kinds of dangers; to say that someone is temperate is to say that if he makes the wrong judgment about what to do, it is not because he tends to overindulge his appetites, and so on.

2.8 But it does not immediately follow that Aristotelian virtues are constitutive standards in my sense. For on Aristotle’s conception, a person’s vices are also constitutive of her will. Recall that according to the constitution requirement, meeting the normative standard in question is what constitutes an entity as the kind of entity that it is. In the Kantian story, that means that an action that is not efficacious and autonomous falls somewhat short of being an action, and the agent who performs it falls somewhat short of being an agent. In *Self-Constitution*, I argue that for a movement to be an action it must spring from the agent as a unified being, as a whole, not just from some force at work in the agent.<sup>19</sup> A heteronomous action, in Kant’s sense, falls short of being an action because it is caused by something at work in the agent – an unendorsed desire, for example – rather than by the agent herself. To make Aristotle a constitutivist in the same sense that I think Kant is, we would have to get the conclusion that somehow actions motivated by a normative conception of the world informed by vice are not fully actions, because they are not fully the agent’s own.

There is certainly material in the *Nicomachean Ethics* one could use to make such a case. There are suggestions throughout the text, but especially in Book 9, that the vicious person is not unified in the way that the virtuous person is. In section 4 of Book 9, Aristotle tells us that the good person’s “opinions are harmonious, and he desires the

same things with all his soul ... And he grieves and rejoices, more than any other, with himself.” (Aristotle 1984a, 9.4, 1166a12–27) “Whereas inferior people... are at variance with themselves, and have appetites for some things and wishes for others.” (Aristotle 1984a, 9.4, 1166b6–7) And bad people “do not rejoice or grieve with themselves, for their soul is rent by faction; ... [they] are laden with regrets.” (Aristotle 1984a, 9.4, 116b18–24) Unfortunately, however, there are also passages that pull against this interpretation, mostly in Book 7, in which Aristotle claims that vice, in contrast with weakness of will or *akrasia*, “is unconscious of itself”<sup>20</sup> and that the vicious person, because he believes himself good, has no regrets.<sup>21</sup> But if we can find a way to argue for the Book 9 conception without losing the contrast with *akrasia*, and so to argue that the Aristotelian vicious agent is not unified, then she can never get fully behind her action in the way that I claim is necessary to make it her own. So I think the case could be made.

But in any case, the view that the state of one’s character is constitutive of one’s will requires a quite specific moral psychology. It requires a moral psychology that supports the view that an agent’s character provides her with a kind of moral perception which in turn informs a normative conception of the world which is then at work in her choices and serves as the moving principle of her action. So far as I can see, an Aristotelian constitutivism, or indeed, anything apart from sentimentalism that might reasonably be called a “virtue ethic,” would have to be based on a moral psychology of that kind.

### 3. Responsible agency and the self-Constitution requirement

3.1 There is one feature of action, however, that I think can only be successfully captured by a Kantian account: that is, it can only be captured if action is governed by standards that not only meet the constitution requirement, but that also meet the self-constitution requirement. The principles governing agency must be such that it is by conforming to them that we *make ourselves into* agents. In this section I will explain why that is important.

In “The Normative Conception of Agency,” I contrasted two accounts of agency, a naturalistic one and a normative one.<sup>22</sup> We might describe them, a little roughly, this way:

NATURALISTIC CONCEPTION OF AGENCY: An person acts when her movements are caused by certain mental states.<sup>23</sup>

NORMATIVE CONCEPTION OF AGENCY: A person acts when her movements are attributable to her *self*.

We can see why the second conception is normative when we ask what it means to say that someone’s movements are attributable to herself – that is, to her *self*. This kind of attribution has two related aspects.

I call the first the “personal efficacy” aspect. The agent’s own efficacy is supposed to be implicated in her agency. Attributing an action to someone is supposed to be different from attributing either a property or an involuntary movement to her in an important way. To regard someone as an agent is to regard her as the one who makes something happen. The efficacy of agency is the agent’s *own* efficacy, we might say, not just that of some chain of causes running through her. It is not immediately obvious that this feature of the concept of agency can be captured by the naturalistic conception of agency. The reason is familiar to us from the old debates about freedom of the will – the mental state that causes the action *itself* presumably has prior causes, so it is not clear why *its* operation should especially represent the agent’s own efficacy, any more than anything else.

I call the second implication the “personal identity” aspect. To say that someone *did* something is to mention what is, in general, a proper ground for responses to her that are personal and normative. Characteristically, we hold adult human beings responsible for their actions; more generally, we take people’s actions to be legitimate or appropriate grounds for responses like love and hate, like and dislike, gratitude and resentment – responses that are focused on the agent herself, and involve taking up an evaluative attitude towards her. These responses show that we take a person’s actions to reflect something essential about *her*, to *represent* her in some way. The 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 whether the legitimacy or appropriateness of these reactions can be explained by the naturalistic conception of agency. We can certainly ask why exactly causation by a certain mental state should count as an expression of the agent’s essential identity. Our mental states might reflect deep and important things about our identities, but they also might not.<sup>24</sup>

I call these two things “aspects” of the normative conception of agency, rather than separate ideas, because in a loose way they view the same fact from first-person perspective on the one hand and the second- or third- person perspective on the other. That fact is the role of the self in action. It is above all when we ourselves act that it seems to us as if we are efficacious – that we ourselves are the ones who are making things happen. And it is when we respond to the actions of others that we ask ourselves what those actions say about their identities, about what sort of people they are. The two aspects are not separate, because both reflect ways in which the idea of an action is supposed to be the idea of a movement, or of the effecting of a change, which has its ultimate *source* in the agent himself, that is, in his essential identity or self. The concept of *responsibility* is the linchpin that unites the two aspects: to apply that concept we need both the idea that the action represents the agent’s own efficacy and that it expresses his essential identity.

Modern philosophers tend to suppose that the naturalistic conception of agency can capture these two aspects of the normative conception of agency if we can just identify the right kind of mental state to be the cause of the action. Earlier philosophers, of course, tried to ground both of these aspects of agency in a metaphysical story about the freedom of the will, sometimes understood in terms of the absence of any causation governing the decision process. The absence of causation was supposed to be what gave the agent the opportunity to insert his own efficacy, and so his own self, into the picture. Such stories seemed outrageous to philosophers with naturalistic inclinations. But they also seemed, most notably to Hume, to pit the two aspects of the normative conception against each other.<sup>25</sup> If the idea that the action involves the agent’s own efficacy rather than that of some cause working through him requires us to regard his actions as uncaused, then *ipso facto* it requires us to regard his actions as uncaused by any features of his identity or self, and so renders the personal identity aspect mysterious. How can the action reflect anything about the agent if nothing about the agent causes it?

3.2 Aristotle himself ran into a version of this problem, and his attempt to deal with it is one of his rare fumbles. On his view, I have already suggested, the moving principle of your action is your normative conception of the world, which is informed by the overall state of your character, at work in your perception of this particular situation. If we think of this as a variant of the naturalistic conception of action, that is what plays the role of the relevant mental state. This account does a good job of explaining what I have called the personal identity aspect of agency. Earlier I pointed out that it is not clear why being caused by an agent’s mental state makes an action the expression of the agent’s essential identity or self. Aristotle has an excellent response to that question: he thinks the mental state in question – the moving

principle of our actions – is a normative conception of the world that reflects our characters. Obviously our characters do express something deep and intimate about our “selves.” If the normative conception of the world that is informed by our characters is the moving principle of our actions, then insofar as we are agents, we are *identical* with our characters. So of course our actions express something normatively important about ourselves.

But this very success gives Aristotle trouble with explaining the other aspect of action. Before I noticed that we might wonder how the naturalistic conception of action can capture the personal efficacy aspect of the idea of action: after all, the mental state itself presumably has prior causes, so it seems to be the efficacy of those prior causes that is at work. In the same way, we can ask Aristotle what makes an action being caused by one’s normative conception of the world an expression of one’s own personal efficacy, if one’s normative conception of the world is an expression of the state of one’s character. After all, the state of one’s character itself has presumably has prior causes, and so it might seem to be their efficacy that is at work. Aristotle himself imagines the objection:

Now some one may say that all men aim at the apparent good, but have no control over how things appear to them but the end appears to each man in a form answering to his character.<sup>26</sup>

Aristotle’s response depends on argument that he has given just before: he thinks he has established that we are “somehow responsible” for our own characters, and therefore for the way things appear to us.<sup>27</sup> A person who is drunk does not know what he is doing, but we hold him responsible anyway, because he got himself drunk. In the same way, Aristotle wants to argue, people are moved inevitably by the state of their characters, but we hold them responsible anyway, because they got themselves into those states of character:

Still they are themselves by their slack lives responsible for becoming men of that kind, and men are themselves responsible for being unjust or self-indulgent, in that they cheat or spend their time in drinking bouts and the like; for it is activities exercised on particular objects that make the corresponding character. (Aristotle 1984a, 3.5, 1114a3–7)

The interlocutor in Aristotle’s mind must have whispered that this answer depends on the idea that the unjust or self-indulgent man *knows* that his actions are forming his character, for in what seems like a spurt of impatience he adds:

Now not to know that it is from the exercise of activities on particular objects that states of character are produced is the mark of a thoroughly senseless person. (Aristotle 1984a, 3.5, 1114b 8–10)

The response is hopeless, of course, since it only raises the iterated question whether someone is responsible for being “thoroughly senseless.” But it might be hopeless anyway, since the unjust or self-indulgent person would not only have to know that the states of his character are produced by his actions, but also that the states he was producing were the wrong ones. But as I noted earlier, Aristotle often (though not always) emphasizes that vice (unlike *akrasia*) “is unconscious of itself,”<sup>28</sup> and the vicious person thinks that he is doing what he ought to do.<sup>29</sup>

3.3 How does the idea of self-constitution explain the two aspects of agency? The Aristotelian view shares with the naturalistic conception of agency the idea that an action is expressive of your own identity if it arises from a certain kind of mental or psychological cause. Aristotle’s conception is better than the purely naturalist conception because the specific mental cause he identifies – your normative conception of the world – is one that necessarily embodies and expresses your character. The self-constitucionalist, however, rejects the idea that what makes an action expressive of yourself is the kind of

mental cause from which it springs. Instead the self-constitucionalist holds that you *constitute* an action as your own and so as expressive of yourself by the way that you act, specifically by the principles upon which you act.

Start with the personal efficacy aspect. To be personally efficacious – to be the one who makes things happen – is to be free; in Kant’s sense, spontaneous. Kant argues that to be free we would have to be autonomous, to act from laws that we impose on our own conduct. This is what the categorical imperative instructs us to do. Since the laws in question are supposed to 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 to be personally efficacious. In a sense, they tell us *how* to be efficacious. Following the hypothetical imperative is what makes our actions efficacious and following the categorical imperative is what makes that efficacy our own. In other words, the idea of personal efficacy is captured by the combined *content* of these two Kantian norms. That means that the way we constitute our personal efficacy is by following those norms.

Think of it this way. In *Self-Constitution*, I argued that 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 as 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 free or personally efficacious being would do. That is, suppose we could work out *which* laws such a being would follow, as Kant supposes we can. That is the view that applying 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 *just is* a conception of yourself as bound by those laws. Then you would be personally efficacious. 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 free being would, then you are free. This is what Kant has in mind when he says: “Now I say: every being that cannot act otherwise than under the idea of freedom is actually free, in a practical respect, precisely because of that.” (Kant 1998, 4, 448) No doubt there is some causal story about how you arrived at all these thoughts, but that does not matter. It is the *formal content* of the principles we follow, not the efficient-causal story about how we came to follow them, that explains how we can be personally efficacious.<sup>30</sup> By following norms with this formal content, you *take* control of your own movements. You make yourself free.

Now turn to the personal identity aspect of agency. The categorical and hypothetical imperatives are formal principles in the sense that they give our actions the *form* of personal 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 material content. The material content of our principles determines what kind of causes we make ourselves into. In the course of choosing our actions, we constitute our own practical identities, as I have elsewhere called them.<sup>31</sup> We make ourselves into a friend, teacher, colleague, benefactor, or whatever, by imposing the form of personal efficacy on principles derived from those roles. With this idea in place, we get the conclusion that Aristotle wanted: we make ourselves into the people who we are, we are responsible for who we are. So the selves whom we are rendering personally efficacious, if I may put it that way, are the possessors of those very practical identities, or if you prefer the Aristotelian story, of those characters. 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.

So the idea is that by choosing our actions in accordance with certain principles, 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. It is appropriate because these are essential features of yourself, but features that you bring into existence through your own personal efficacy.

If these arguments work, the idea that actions are expressions of your own efficacy and at the same time expressive of your “self” are supported by the idea of self-constitution. The naturalist’s mistake, and Aristotle’s mistake, is to try to explain the normative aspects of agency in terms of the kind of mental cause that leads to action rather than in terms of the content of the formal principles that govern action. We make ourselves into agents by following principles that express the formal essence of personal efficacy, and we make ourselves into the particular agents who we are by the material content on which we impose those formal principles.

#### 4. The role of the virtues

4.1 Aristotle’s conception of the moral virtues is a conception of what it means for our passive or receptive capacities to be in a good state for moral action. If we take this one way, as I myself have done in the past, this is a conception of the virtues anyone can accept, regardless of what sort of moral theory she holds, and whether or not she champions a form of “virtue ethics.”<sup>32</sup> No matter how good an agent’s principles are, she will not act on them reliably or correctly unless she sees the world in a certain way, and the state of what Aristotle calls her “character” – the state of her passions and emotions – does play a role in determining how she sees the world, normatively speaking. The most carefully reasoned commitment to the principle of helping others for its own sake, for example, will not lead you to practice beneficence when you should if the fact that someone is in trouble or need does not strike you as an occasion when the question of giving help should at least come up. And no doubt you will be liable to attacks of self-deception if helping always presents itself to you as a disagreeable chore. As Kant himself says:

But what is done not with pleasure but merely as compulsory service has no inner worth for one who attends to his duty in this way and such service is not loved by him; instead he shirks, as much as possible occasions for practicing virtue. (Kant 1996, 6, 484)

On this interpretation, when we say that someone has a virtue, we are not merely gesturing third-personally at the fact that she has good principles. We are saying that her passions and emotions are in a condition that is conducive to doing the right thing, to acting on the right principles. But we can say that while thinking of the will or at any rate of the agent’s principles as something that is separate from virtues themselves.

But that is not how Aristotle himself thought of the virtues. As we have seen, he thinks of them as constitutive of a person’s will. One way to describe the difference between Kant and Aristotle is say that Aristotle seems to think we must articulate our principles (to the extent that we can) by extracting them from our normative conception of the world, while Kant thinks that an *a priori* thought about the form of the principles identifies the right ones. In fact, I believe that Aristotle thinks we arrive at ethical principles from our normative conception of the world by an act of *nous* or rational apprehension, in a way that parallels the way he thinks we are able to extract the forms of things from perception.<sup>33</sup> The process is one that looks like induction, but enables us to arrive at necessary truths.

I do not think a Kantian can believe in such a power without believing in a kind of harmony between the mind and the world that Kantians do not suppose exists, or that we could ever know exists. Nevertheless, there is a way that a Kantian could take on board a conception of the role of the virtues that is closer to Aristotle's own view. We could understand Aristotelian virtues as essential to good judgment, in a specific sense of "judgment" – the ability to apply a principle.

As Kant himself urges:

A physician, a judge, or a ruler may have at hand many excellent pathological, legal, or political rules, even to the degree that he may become a profound teacher of them, and yet, nonetheless, may easily stumble in their application. For, although admirable in understanding, he may be wanting in natural power of judgment. He may comprehend the universal in abstraction, and yet not be able to distinguish whether a concrete case comes under it. Or the error may be due to his not having received, through examples and actual practice, adequate training for this particular act of judgment.<sup>34</sup>

We can chase down the reasons why an action is right by making our principles increasingly specific, but at some level, the ability to articulate why an action is right comes to an end, and we just have to *see* that the case at hand really does fall under the principle. So there is a role for normative perception even in a theory that emphasizes principles.

With this thought in hand, we might almost begin to see the difference between Kant and Aristotle as an argument over a matter of degree. Aristotle tends to underplay the possibility of articulating moral principles, while Kant sometimes seems to overplay the ease of applying them once we have used the categorical imperative to get them articulated.<sup>35</sup> Articulate principles and something akin to a kind of skill in applying them are both needed, we might say, for reasoning to action. In fact presumably the more virtuous you are, the better you will be at *articulating* the reasons why the action is right. After all, the difference between merely natural virtue and fully moral virtue in Aristotle's account is this:

For it is not merely the state in accordance with the right reason, but the state that implies the presence of the right reason, that is virtue, and practical wisdom is right reason about such matters. (See Aristotle 1984 6.13, 1144b26–28)

In a theory that combined a Kantian theory of principle with an Aristotelian account of judgment, if such a theory proved possible, the moving principle that makes the action attributable to an agent would, in the case of a genuinely virtuous agent, not just be the thought that she ought to do a certain thing for certain reasons. It would be that thought lit up, so to speak, made meaningful, made *perceptible*, by a normative conception of the world that at once embodied that principle and would be reflected in the agent's character.

### Disclosure statement

No potential conflict of interest was reported by the author.

### Notes

1. It is because of my appeal to the constitution requirement that I identified my view in *Self-Constitution* as a form of Platonism. To say that an object fits under a certain concept only to the extent that it conforms to the constitutive standards given by that concept is essentially the same as to say that the object gets its identity from its participation in a form that represents a perfect version of that kind of object.

2. In earlier work I treated “internal” and “constitutive” as more or less synonymous, but in this essay I am using “internal” to name a broader category of standards.
3. See Plato 1997, 353 b–c in the Stephanus numbers found in the margins of most translations; Aristotle 1984a, 2.6, 1106a14ff; *NE* 6.2, 1139a18. References to Aristotle will be given using the standard Bekker page, column and line numbers.
4. I am using the word “principles” in a way that ranges both over intentions and Kantian maxims.
5. Hume 1978, Part 3, Book 1, Part 2; 477–478.
6. Foot 2001, 12. I have added the comparisons to Hume.
7. The everything else that has to be equal is that there is no prudential or moral cost to these actions that would make it seem to me as if they were not worth doing.
8. See Kant 1998, section I and my analysis of the argument in Korsgaard 1996b, 43–76 and Korsgaard 2008d, 176–187. References to Kant’s *Groundwork* and *The Metaphysics of Morals* will be given by the page numbers of the relevant volume of *Kants gesammelte Schriften*, which appear in the margins of most translations.
9. Kant 1998; 4, 398–399.
10. See Kant 1998, 4, 399–401. Kant creates confusion by using the word “maxim” in two different ways. Sometimes he uses it to refer to what we might call a first-order maxim, doing a certain act for a certain end, which we adopt on further grounds. When he gives examples in *Groundwork* II of agents who ask themselves whether their maxims can serve as universal laws, he is using the term that way. At other times he uses it to refer to an agent’s total motivational state, a maxim along with the grounds on which it has been chosen. That’s how he’s using it when he says, in the passages referred to, that the difference between naturally sympathetic action and beneficent action from duty does not rest “in the purpose to be attained by the action” but “in the maxim in accordance with which it is decided upon.” (Kant 1998; 4, 399).
11. See Hume 1978, 3.1.1.
12. For further discussion, see Korsgaard 2009b.
13. See Foot 2001, 27–37. She is drawing on Thompson 1995.
14. For defense of the view that *ergon* means how something does what it does, see Korsgaard 2008b.
15. See Aristotle 1984a, 1.7 1097b21–1098a17.
16. There is a parallel in the theoretical case. Any agent – any animal in Aristotle’s sense – must form some sort of a representation of the environment. Non-rational animals do that through perception. Rational animals do it by constructing the demonstrations that constitute scientific knowledge, *episteme*.
17. For an earlier defense of this reading, see Korsgaard 2008c.
18. See Aristotle 1984a, 6.13, 1144b30–1145a2.
19. See Korsgaard 2009a, 1.4.1.
20. See Aristotle 1984a, 7.8, 1150b35–36.
21. See Aristotle 1984a, 7.9 1152a4–6; see also 7.8 1151a11–13 and 7.8 1151a6–7.
22. In this part of the paper I draw heavily on the parallel discussion in Korsgaard 2014, although I have changed some of the terminology.
23. This formulation needs modification, at least for people (there might not be much more to it for simple animals), because of the problem of deviant causal chains. In Davidson’s famous example, a climber holding another climber by a rope has the thought that he could make his own position more secure by dropping the rope, and finds the thought that he might do such thing so unnerving that he drops the rope. Kant defines agency as the capacity to be, by means of one’s representations, the cause of the object of those representations. (See Kant 1996, 6, 211) In this case, the climber’s representation (“I could just drop the rope ...”) is the cause of the object of his representation (he drops the rope), yet this is not something that *he* intentionally does – in Kant’s terms, *he* is not the cause. I leave aside this problem for now.
24. I once asked Michael Bratman whether an intention produced by a post-hypnotic suggestion would be just as much the person’s own intention as any other, and he said yes. For further discussion see Korsgaard 2014, 200–202.
25. See Hume 1978, 2.3.3, especially 410–412.
26. Aristotle, 1984a, 3.5, 1114a31–1114b3. The translation has “him” rather than “them.”
27. See Aristotle, 1984a, 3.5, 1142b2.
28. See Aristotle, 1984a, 7.8, 1150b 35–36.

29. See the references in note 21.
30. The conception of agency that I am describing here does not support as “deep” a conception of moral responsibility as some of the traditional defenders of free will hoped to defend. It would not, for instance, justify an omnipotent God in consigning a bad person to eternal damnation. For that, it would matter how we came to follow the principles that make us efficacious. I do not regard this as a loss.
31. For the notion of practical identity, see Korsgaard 1996a, lecture 3; and Korsgaard 2009a, 1.4, pp. 18–26.
32. See for instance Korsgaard 2008d.
33. See Aristotle 1984a, Book 6, Section 11.
34. See Kant 1965, 178. A 134/B173 in the pagination of the A and B editions found in the margins of most translations.
35. See, for instance, the discussion at Kant 1998, 4, 403–404.

### Notes on contributor

*Christine M. Korsgaard* is the Arthur Kingsley Porter Professor of Philosophy at Harvard. She is the author of *The Sources of Normativity* (Cambridge, 1996), *Creating the Kingdom of Ends* (Cambridge, 1996), *The Constitution of Agency* (Oxford, 2008), *Self-Constitution: Agency, Identity, and Integrity* (Oxford, 2009), and *Fellow Creatures: Our Obligations to the Other Animals* (Oxford, 2018).

### References

- Aristotle. 1984a. “The *Nicomachean Ethics*. Translated by W. D. Ross and revised by J. O. Urmson.” In *The Complete Works of Aristotle: The Revised Oxford Translation*. Vol. I., edited by Jonathan Barnes, 1728–1867. Princeton: Princeton University Press. References to Aristotle’s works are given by the standard Bekker page, column, and line numbers. NOTE: In quoting from this edition I have deviated from the translation in two ways: I translate “*ergon*” as “function” rather than “work,” and “*arete*” as virtue rather than “excellence”.
- Aristotle. 1984b. “On the Soul. Translated by J. A. Smith in.” In *The Complete Works of Aristotle: The Revised Oxford Translation*. Vol. I., edited by Jonathan Barnes, 641–692. Princeton: Princeton University Press.
- Foot, Philippa. 2001. *Natural Goodness*. Oxford: Clarendon Press.
- Hume, David. 1978. *A Treatise of Human Nature*. 2nd ed., edited by L. A. Selby-Bigge and revised by P. H. Nidditch. Oxford: Clarendon Press.
- Kant, Immanuel. 1965. *Critique of Pure Reason*. Translated by Norman Kemp Smith. New York: Macmillan, St Martin’s Press.
- Kant, Immanuel. 1996. *The Metaphysics of Morals*. Translated by Mary Gregor. Cambridge: Cambridge University Press.
- Kant, Immanuel. 1998. *Groundwork of the Metaphysics of Morals*. Translated by Mary Grego. Cambridge: Cambridge University Press.
- Korsgaard, Christine M. 1996a. *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Korsgaard, Christine M. 1996b. “Kant’s Analysis of Obligation: The Argument of *Groundwork I*.” In *Creating the Kingdom of Ends*, 43–76. Cambridge: Cambridge University Press.
- Korsgaard, Christine M. 2008a. “The Normativity of Instrumental Reason.” In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, 27–68. Oxford: Oxford University Press.
- Korsgaard, Christine M. 2008b. “Aristotle’s Function Argument.” In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, 129–150. Oxford: Oxford University Press.
- Korsgaard, Christine M. 2008c. “Aristotle on Function and Virtue.” In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, 151–173. Oxford: Oxford University Press.
- Korsgaard, Christine M. 2008d. “From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action.” In *The Constitution of Agency*, 174–206. Oxford: Oxford University Press.
- Korsgaard, Christine M. 2009a. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Korsgaard, Christine M. 2009b. “Natural Motives and the Motive of Duty: Hume and Kant on our Duties to Others.” *Contemporary Readings in Law and Social Justice* 1 (2): 8–35.

- Korsgaard, Christine M. 2014. "The Normative Conception of Agency." In *Rational and Social Agency: The Philosophy of Michael Bratman*, edited by Manuel Vargas and Gideon Yaffe, 190–214. Oxford: Oxford University Press.
- Plato. 1997. "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.
- Thompson, Michael. 1995. "The Representation of Life." In *Virtues and Reasons*, edited by Rosalind Hurthhouse, Gavin Lawrence, and Warren Quinn, 25–64. Oxford: Oxford University Press. later part 1 of Thompson, Michael, *Life and Action: Elementary Structures of Practice and Practical Thought*. Cambridge, MA: Harvard University Press, 2008.