20

CONSTITUTIVISM

PAUL KATSAFANAS

Constitutivism is the view that we can justify fundamental normative claims by showing that agents become committed to these claims merely in virtue of acting.[1] Constitutivists aspire to show that action has structural features – constitutive aims, principles, or standards – that are present in each instance of action and that generate substantive normative conclusions. In showing that the authority of fundamental normative claims is sourced in our own actions, constitutivists hope to avoid familiar objections to justificatory projects in ethics.

This chapter provides a very brief overview of constitutivism. The first section outlines the basic structure of constitutivism. The second and third sections examine how the constitutive feature would generate normative results. The fourth considers an objection to the constitutivist theory. The fifth distinguishes between constitutivist theories that attempt to provide fully general accounts of normativity and more modest versions. The sixth section asks how much normative content constitutivist theories are supposed to generate. The final section concludes.

THE BASIC STRUCTURE OF CONSTITUTIVISM

First, a word on constitutivism's origin. Constitutivism can be seen as emerging in response to concerns about the relationship between normative claims and facts about the agents to whom they apply. Beginning in the mid-seventies, this was one of the focal points for writings on ethics. J. L. Mackie's influential argument from queerness prompted some of these debates. Mackie wrote: "an objective good would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has *to-be-pursuedness* somehow built into it" (Mackie 1977: 40). It's mysterious what this property could be.

[1] See Alvarez and Hyman, this volume, for discussion of the history of philosophy of action.

On the basis of these sorts of reflections, Bernard Williams (1981a) argued that an agent has a reason to *X* only if there is a "sound deliberative route" from the agent's subjective motivational set to *X*-ing. These claims spawned an enormous literature, with philosophers falling into two broad camps: externalists claim, and internalists deny, the following claim: "An agent *A* can have a reason to *X* even if *A* does not have, and would not have after procedurally rational deliberation, a motive whose fulfillment would be promoted by *X*-ing." Each side of the debate has costs. Briefly, internalists have difficulty establishing genuinely universal normative claims, such as "you have reason not to murder." After all, it seems possible for an agent to lack any motives that are suitably connected to not murdering. Externalism vouchsafes universal normative claims, makes it less obvious how these claims connect to motivation: By hypothesis, some of them will be entirely disconnected from the agent's actual motives.[2] This can make external reasons look, as Mackie puts it, decidedly queer.

In response to these debates, externalists and internalists attempted to diagnose problems with one another's views. Attempts at synthesis emerged, with Michael Smith (1994) arguing that we should reconfigure the debate in terms of what an idealized, perfectly rational version of the self would desire, and John McDowell urging a focus on the *phronimos* (McDowell 1995b). Meanwhile, neo-Kantian theories developed by Barbara Herman (1996), Christine Korsgaard (1996b), and others tried to reconcile universal and categorical demands with the idea that these demands issue from the agent herself.

Constitutivism can be seen as a new entry in these well-worn debates. For constitutivism operates by showing that action or agency has features that generate universal normative commitments. Although the details vary, there are two general strategies: an *aim-based* version of constitutivism and a *principle-based* version. The aim-based version tries to show that there is an aim present in every episode of action, and then argues that the aim generates normative reasons; the principle-based version tries to show that each action is governed by a particular normative principle. Let me explain.

I'll begin with the aim-based version. We can define constitutive aims as follows:

(Constitutive Aim) Let A be a type of attitude or event. Let G be a goal. A constitutively aims at G iff

(i) each token of A aims at G, and
(ii) aiming at G is part of what constitutes an attitude or event as a token of A.

[2] I discuss these points in more detail in Katsafanas 2013.

The clearest examples of activities with constitutive aims are games. Take chess. Arguably, chess has the constitutive aim of checkmate: Every token of chess-playing aims at checkmate, and aiming at checkmate is part of what constitutes a token action as an episode of chess-playing.[3] In other words, if you're moving chess pieces about on a board without aiming at checkmate, you're not playing chess; and if you are moving pieces about on a board while aiming at checkmate, this is part of what makes your movements count as episodes of chess-playing.

The constitutivist about action hopes to show that action itself has a constitutive aim. This is, of course, counterintuitive. Why think that action has any constitutive features whatsoever? It's clear enough the activities such as chess have constitutive features. But that's not surprising: Chess is a game, defined by its rules, with clear criteria of success and failure. How could we show that action itself has constitutive features?

David Velleman tries to show that action has a constitutive aim of self-understanding: In each case of action, the agent aims to attain understanding of what she is doing and why (Velleman 2009). Christine Korsgaard argues that action has a constitutive principle of self-constitution, where self-constitution is secured by acting on the Categorical Imperative (Korsgaard 2009). And I argue that action has two constitutive aims. First, action constitutively aims at a form of reflective approval: In deliberative action, I aim to perform actions of which I approve, where this approval is stable upon the revelation of further information about the way in which the action is motivated. Second, action constitutively aims at challenge-seeking, in the sense that each episode of action aims not merely at some determinate goal, but also at the encountering and overcoming of challenges in the pursuit of that goal (Katsafanas 2013).

Of course, each of these arguments is controversial: You wouldn't just glance at action and conclude that it aims at self-understanding, for example. So each theory requires a foundation in an account of intentional action that is independently plausible. If we just assert, for example, that action is movement governed by the Categorical Imperative, then Korsgaard's account would follow. But this would be of no interest; all the work would be done by the initial assertion. Thus, constitutivists typically aspire to begin with some very minimal, widely accepted account of intentional action, and show that this minimal account entails that action has a constitutive feature.

[3] Actually, this is a bit of a simplification: You can also play chess while aiming to attain a draw. So, to be precise, we should say that the constitutive aim of chess-playing is *attaining checkmate or a draw*. But, in order to avoid clunky formulations, I'll ignore this complication above. It does not affect any of the arguments.

The most powerful version of constitutivism would start with an absolutely uncontroversial account of action and show that it yields a constitutive feature. For example, suppose we can start with the idea that action is goal-directed movement, and show that this entails that action has a constitutive aim of self-constitution. That would be of enormous interest; nearly everyone accepts the initial account of action, so, if the arguments work, it would follow that everyone should accept the constitutivist's conclusions. But suppose, by contrast, that we must start with a more substantive and controversial account of action. Then, even if the constitutivist's arguments work, they will only be as powerful as the arguments for the initial theory of action. For example, if Korsgaard's constitutivism starts with the assumption that action aims at governance by the Categorical Imperative, then all the work is done by the initial defense of the conception of action. (I've argued elsewhere that something like this is true: Both Velleman and Korsgaard present themselves as starting with minimal and almost universally accepted accounts of action, but, in fact, their arguments equivocate; they end up relying on substantive, highly controversial, and undermotivated accounts of action.)

So how do the constitutivist arguments begin? In each case, the theories begin with a relatively uncontroversial account of action and then proceed to argue that the uncontroversial account yields surprising conclusions:

(1) Korsgaard begins by claiming that action is simply movement attributable to the whole agent, rather than to some part of the agent (2009: 18). She then argues that an action is attributable to a whole agent only if the agent acts on a normative principle (119–20); that different normative principles engender different degrees of agential unity (120–79); that the Categorical Imperative is the only principle that fully unifies us (169–77); and, finally, that it follows that the Categorical Imperative is the constitutive principle of agency.

(2) Velleman begins by claiming that action is immediately known, in the sense that agents have non-observational knowledge of what they are doing and why they are doing it (Velleman 1989). He then argues that we can best account for the presence of this immediate knowledge by positing that agents have a standing desire for self-understanding, which inclines them to act in ways that they antecedently expect to act (as he puts it, "the agent attains contemporaneous knowledge of his actions by attaining anticipatory knowledge of them" [2004: 277]). On this basis, he argues that action constitutively aims at self-understanding.

(3) I begin with two claims: that we aim to perform actions of which we approve and that we aim not only to attain end states, but also to manifest particular forms of activity. The latter claim is based on a fact about human

motivation: In addition to desires for particular objects, we are motivated by drives. Drives are characterized by their aims: The sex drive aims at sexual activity, the aggressive drive at aggressive activity. Drives incline us to seek objects upon which to vent this activity, but the objects can be adventitious; the essential thing is the expression of the aim. I argue that drives operating in this way, continuously generating objects of desire, can be understood as inducing an aim of seeking to encounter and overcome certain types of resistances. Moreover, I argue that in deliberative action, we take the outcome of deliberation to settle what we are going to do; and this involves assessing an outcome as preferable to others; and this involves aspiring to approve of what you do; and, this involves seeking for the approval to be stable in the face of further information about the motivational inputs to our deliberation. From this, I argue, we can derive a constitutive aim of agential activity. Coupling it with the first aim, we get the following structure: Action constitutively aims at selecting activities such that they involve overcoming challenges that induce resistances; and, in light of the first aim, we're committed to approving of this aim, as something that structures all of our actions. (See Katsafanas 2013 for the details.)

These are just sketches and probably raise as many questions as they answer. But my hope is that they give some indication of how the constitutivist argues: We start with an independently plausible account of action and try to show that it yields substantive normative conclusions.

THE REASONS GENERATED BY THE CONSTITUTIVE FEATURE

Suppose we can in fact show that action has a constitutive aim. What would follow from this? The fact that action has some constitutive aim is a merely descriptive premise; how do we get normativity from this?

This brings us to the second step of the constitutivist project: showing how constitutive features generate reasons. I've suggested that the constitutivist should defend a principle of the following form:

(Success) If X aims at G, then G is a standard of success for X, such that G generates normative reasons for action.

Returning to the chess example: If chess has a constitutive aim of checkmate, then by Success players will be committed to treating checkmate as providing them with reasons for action. Success should be understood as a fully general claim about aims: If you have an aim, then all else being equal you have normative reasons to fulfill it. For example, if Nolan aims at eating cake and

Franny doesn't, then all else being equal Nolan has a reason that Franny lacks: a reason to eat cake. (I'll consider objections below.)

If we accept something like Success, then reasons are fairly cheap: Any aim will generate prima facie reasons. What's interesting about the reasons generated by the constitutive aim, though, is that they would have universal scope. They would apply to agents just in virtue of the fact that they are performing actions. Just as the chess player's reason to promote checkmate arises from the nature of the game and thus applies independently of his contingent psychological preferences, so too the agent's reason to realize action's constitutive aim would arise from the nature of agency as such and would thus apply independently of her contingent psychological states. So the reasons derived from the constitutive aim would be universal. Many philosophers regard this as a criterion of adequacy for ethical theories.

But should we accept something like Success? Although Velleman (2009) and Katsafanas (2013) defend versions of it, it's not entirely uncontroversial. After all, we might deny that aims generate reasons. This skepticism is sometimes motivated by reluctance to say that aiming at reprehensible activities, such as murder, provides one with a reason to murder. Accordingly, we might think that aims generate reasons only if we antecedently have reason to adopt the aim; or we might deny any connection whatsoever between aims and reasons. However, I've elsewhere argued that the standard ways of arguing against principles like Success in fact present no difficulties for the constitutivist. While there is not enough space to explain this point fully, I'll give one example. Suppose the objection to Success is based on Broome-style wide-scope readings of the instrumental principle. Although this might seem problematic, it actually presents the constitutivist with no trouble at all. It merely requires a reformulation of Success in something like this form: Rationality requires that if you have an end G, then [either you give up this end or you take the necessary and available means to G]. (See Broome 1999.) Reformulating Success in terms of rational requirements won't bother the constitutivist, as the constitutive aim won't be capable of being abandoned; thus, rationality will require that you take the necessary and available means to attaining it. Analogous arguments are available for other standard interpretations of the relationship between aims, goals, and reasons (see Katsafanas 2013: chapter 2 for the details).

If we're terribly skeptical about these responses, though, a second variant of constitutivism is available. Recall the distinction between aim- and principle-based versions of constitutivism. Rather than arguing that action has a constitutive aim and trying to show that the aim generates reasons, we could directly argue for the presence of a normative feature in action. Christine Korsgaard pursues this strategy. To simplify a bit, suppose we accept a roughly Kantian

account of action according to which action is movement governed by the Categorical Imperative. Then each instance of action would be governed by the Categorical Imperative, and being governed by the Categorical Imperative would be part of what makes an event qualify as an action. So the Categorical Imperative would be the constitutive principle of action. Moreover, we would not need a second step to establish the normativity of this principle: The principle already contains normative content.

I've argued elsewhere that we should see the aim-based version of constitutivism as roughly Humean, and the norm-based version as Kantian (Katsafanas 2013). The aim-based version establishes the universality of a particular aim in action, and then appeals to some independent principle for generating reasons from this aim. The principle-based version argues directly for a normative notion of action. Accordingly, the theories will face objections at different points, as I'll explain below.

THE STATUS OF THE REASONS GENERATED BY THE CONSTITUTIVE FEATURE

I've mentioned one desideratum for an ethical theory: that it provide universal reasons. But some philosophers also seek to show that ethics generates overriding reasons. Suppose action has a constitutive feature (an aim or a principle) that yields normative reasons. Notice that it does not immediately follow that the reasons derived from the constitutive feature are overriding. Return to the chess example. Chess players have reason to checkmate their opponents, but other reasons will also be present: The player may aspire to enjoy the game, let's say, and therefore have reason to do what promotes enjoyment. And she may have reason to try out some new strategy, rather than taking the most direct and obvious route to checkmate. These reasons may lead her to make different choices than she would if they were absent.

So the reasons derived from the constitutive aim interact with the reasons arising from other sources. But the reasons derived from the constitutive aim do have a special status: The constitutive aim can't be abandoned without abandoning the activity. You can't fully disengage from it. You can, however, disengage from the other sources of reasons: You can't give up the aim of checkmate without ceasing to play chess, but you can give up the aim of enjoyment or of trying out a new strategy.

In this sense, the reasons derived from the constitutive aim are inescapable, whereas the reasons derived from other sources are escapable. Thus, constitutivism generates an inescapable normative standard while allowing all other standards to be escapable.

Of course, the constitutive feature is escapable in one sense: We can stop participating in the activity governed by the constitutive aim. I can stop playing chess, for example, and thereby escape any reasons generated by chess's constitutive features. But notice that action itself is crucially different. Suppose a constitutivist can establish that action itself has some constitutive aim. Certainly, agents could escape this aim by ceasing to perform intentional actions. They can go to sleep, for example. But, given that ethical theory aspires to provide norms for intentional action, constitutivism provides reasons with a scope as wide as that of ethical theory. The reasons derived from the constitutive aim of action would be inescapable for those performing intentional actions.

But it still does not follow that the reasons derived from the constitutive aim are weightier than other reasons. Nor does it follow that the reasons generated by the constitutive feature are particularly substantive. In principle, it's possible that constitutivism generates only weak and always overridden reasons. So we'll have to say more about the status of these reasons. Some constitutivists, like Korsgaard, want to show that constitutivism justifies most of the content of modern morality; others, like Velleman and I, think it can't do *that*, but can yield a number of interesting normative results. I'll explore these possibilities below.

CAN WE BE ALIENATED FROM THE CONSTITUTIVE FEATURE?

But there's also a different reason for worrying about the attempt to anchor normativity in inescapable features of agency. After all, can't you be moved by the constitutive aim, and regard it as inescapable, while being alienated from it? It certainly is possible to deplore aims that figure in wide swathes of agency: Consider the way in which religious ascetics respond to desires for material comfort, sexual pleasure, and so forth. They may regard these aims as ineradicable, but nonetheless contend against them. The constitutive aim would be a bit different: Even though desires for comfort and sex are widespread, they are not omnipresent; they are not aims that are manifest in every episode of action. So the person who deplored the constitutive aim would be even more radical than the ascetic. Nevertheless, can't we take the same attitude toward constitutive features, coherently deploring an omnipresent aim?

Some constitutivists recognize this difficulty and try to respond to it. Velleman, for example, claims that the constitutive aim of self-understanding is deliberatively self-validating:

even if we all of us share this aim . . . you can still ask whether we ought to have it, and why. To my ear, this question sounds like a demand for self-understanding . . . So interpreted, the question demands that the constitutive aim of action be justified in relation to the criterion set by the aim itself. Such a justification is easy to give. (Velleman 2009: 138)

In other words, if action constitutively aims at self-understanding, then asking why we ought to have and to fulfill this aim is a request for self-understanding: It is a request that presupposes the very aim that it questions. In that sense, Velleman claims, it is self-validating: Any attempt to question it will already presuppose it. So alienation from the constitutive aim isn't possible (or at least isn't coherent).

I attempt a different answer. As I see it, one of the constitutive aims of agency is approving of one's action while aiming that this approval be stable in the face of further information about the action's etiology. If every episode of action contains a second constitutive feature – call it F – then F will be part of the etiology of each action. Thus, insofar as one meets the first constitutive condition, one will aspire to meet the second: Insofar as one aims to approve of one's action in light of further information about the action's etiology, and insofar as F is part of each action's etiology, the agent will be committed to taking F's presence as not undermining her approval of the action. So the agent is left with two options: Deplore action as such, or approve of the constitutive feature. There are only two fully coherent responses: Total rejection or total acceptance. Or so, at any rate, I argue (Katsafanas 2013).

Regardless of whether these strategies work, the constitutivist does need to say something about the possibility of alienation from or rejection of the constitutive aim.

DOES CONSTITUTIVISM PROVIDE A FULLY GENERAL ACCOUNT OF NORMATIVITY?

Constitutivism operates by identifying a constitutive feature of agency which is then taken to generate substantive normative conclusions. But this is just a schema, identifying a type of ethical theory. There are many ways to instantiate these features.

A central dimension on which constitutivist theories vary concerns their aspirations: The constitutivist can attempt to provide a fully general account of normativity or a more localized account. Korsgaard, for example, tries to offer an account of normativity as such: She writes that

the *only* way to establish the authority of any purported normative principle is to establish that it is constitutive of something to which the person whom it governs is committed – something that she either is doing or has to do . . . The laws of logic govern our thoughts because if we don't follow them we just aren't thinking . . . the laws of practical reason govern our actions because if we don't follow them we just aren't acting. (Korsgaard 2009: 32)

So, according to Korsgaard, both practical and theoretical normativity must be accounted for in terms of constitutive features; anything less would be a failure.

Some philosophers have mistakenly taken this to be an essential component of the constitutivist project. Matthew Silverstein, for example, writes that principles such as Success cannot succeed:

[F]or they draw on the explicitly *normative* principle that one has a reason (or is at least rationally required) to achieve or promote one's aims or ends. The whole point of constitutivism is to derive action's standard of correctness from *non-normative* facts about the nature of agency. Action's constitutive norm is supposed to be the foundation of *all* practical normative authority . . . And so it cannot rely on some other norm for justification. (Silverstein 2016: 233)

Now, Silverstein is right about Korsgaard: As the quotation above demonstrates, she thinks that constitutivism is the only possible justificatory strategy in ethics. But Silverstein is wrong that the "whole point of constitutivism" is to provide a foundation for all normativity. We needn't be committed to this maximally ambitious version of constitutivism; less ambitious versions, which seek to secure *incontestable* sources of practical normative authority, without thereby trying to secure *all* sources, would still yield substantial and distinctive results. After all, we could treat constitutive aims as one source of normativity among others. There would be nothing incoherent about combining constitutivism with realism: We could hold that certain norms arise from constitutive features, whereas others are just irreducible normative truths. Or we could combine constitutivism with a Humean account: We could hold that there is some function from our actual or hypothetical subjective motivational states to reasons, while also endorsing the constitutivist schema. My constitutivist account – and, if I understand it, Velleman's – takes this latter form. The interest of constitutivism would then lie in its ability to generate *inescapable* reasons, rather than (as for Korsgaard) *all* reasons.

It may be helpful to offer a more concrete illustration of the way in which a constitutivist could treat constitutivism as a source of *some*, but not all, reasons. Consider the instrumental principle. This is the best candidate for a constitutive feature of agency: if anything is constitutive of agency, this is. Action is bringing things about. To bring things about requires that you take the necessary and

available means. So, insofar as we are aiming to bring things about, we are aiming at taking the necessary and available means to bringing them about. More precisely:

(1) An agent's A-ing is an action iff in A-ing the agent aims to bring about some end.
(2) An agent aims to bring about an end iff the agent aims to take some of the necessary and available means to this end.
(3) Therefore, an agent's A-ing is an action iff in A-ing the agent aims to take some of the necessary and available means to her end.

From (3), it follows that taking the necessary and available means to one's ends is a constitutive aim of action. If we accept Success, this entails a version of the instrumental principle.

But, Silverstein objects, have we really shown that there's a *reason* for taking the means to your ends? No, he thinks: The mere fact that I inescapably aim at X doesn't yet entail that I have a reason for trying to attain X. To establish that – to move from claims about inescapable aims to claims about reasons – we have to make substantive normative assumptions.

I think here we reach bedrock. One type of philosopher says: Even if you have some inescapable aim, even if the aim is present in everything you do, you can still ask whether you have a reason for pursuing the aim. And another type of philosopher says: That's nonsense. When you ask whether you have a reason for A-ing, you're asking whether you should pursue A. But in the constitutive case, the question is idle. There's no alternative to A-ing. So your question about whether there's a reason to A, although it has the grammatical form of a question, is moot. Inescapability is the fundamental form of normativity. When you're asking about normativity, you're asking whether you should pursue something; an appropriate answer to that question is showing that you cannot do anything but pursue the thing. (For a more detailed treatment of this point, see Katsafanas 2013: chapter 2.)

But suppose we're philosophers of the first type: We insist that we can question whether there's a reason to act on inescapable aims. Then we have to rest content with the idea that constitutivism relies on a further normative principle specifying that there's a reason or a rational requirement to take the means to your ends. Constitutivism so interpreted wouldn't be a fully general account of normativity. But in response, I'm tempted to say: So what? If we can show that our derivation of practical normativity relies on the assumption that we have reason or a rational requirement to take the means to our ends, then we're still in far better shape than other ethical theories. What are worrying and controversial about ethics are questions about substantive goods: We want to know whether we have reason to reject inequality; whether we have reason to

strive for some forms of life and avoid others; whether cruelty is wrong; whether compassion good; and so on. What worries many of us is that these claims might be unjustifiable. But if they're justifiable so long as we accept the claim that we have reason or rational requirements to take the means to our ends, then the concerns are far less pressing. I, for one, would welcome an ethical theory that actually showed that insofar as you're committed to instrumental rationality, you're committed to treating slavery, cruelty, and so forth as wrong. The maximally ambitious, Korsgaardian version of constitutivism has a certain appeal. But it's not all or nothing: Less ambitious, Humean versions of constitutivism still establish substantive and important results.

HOW MUCH CONTENT IS DERIVABLE FROM CONSTITUTIVISM?

I'll close with a related point. Some constitutivists aspire to show that we can derive something like traditional bourgeois morality from the constitutive features of agency; others aspire to show only that *some* substantive norms are derivable. Korsgaard is in the first camp: She claims that we can derive "Enlightenment morality" from the constitutive features of action (Korsgaard 1996b: 123), where "Enlightenment morality" seems to be roughly coextensive with the moral claims typically embraced by contemporary, liberal, highly educated classes in the urban USA and Europe. Velleman denies this, holding instead that we can give a constitutive account of various norms that "favor morality without guaranteeing or requiring it" (2009: 149). My own version of constitutivism is closer to Velleman's than to Korsgaard's: I think we can derive certain principles that enable us to assess competing evaluative and normative claims, but these principles do not result in the justification of a unique ethical view (Katsafanas 2013). So Velleman and I think that constitutivism enables us to rule out some ethical views and to recommend others, but not to reach a justification of a single ethical view.

CONCLUDING REMARKS

I've given a very brief overview of the constitutivist strategy, discussed some ways in which the constitutivist theories vary, and offered clarifications in response to standard objections and confusions. While much work remains to be done, I hope this chapter gives an indication of how constitutivism works and why it is appealing.