

# Constitutivism about practical reasons

Paul Katsafanas

Draft of March 6, 2014

A focal point in recent work on practical reason is the idea that we might ground normative claims in facts about the nature of agency. According to *constitutivism*, certain normative claims apply to us merely in virtue of the fact that we are agents. Proponents of this view argue that there are features of action that both constitute events as actions and yield normative standards of assessment for action. For example, David Velleman has argued that all actions share the common, higher-order aim of self-understanding. The presence of this aim both distinguishes genuine actions from mere events and yields a standard of assessment for action: we have reason to perform those actions that yield self-understanding.

The attractions of constitutivism are considerable. By anchoring normativity in necessary features of agency, constitutivism provides a way of justifying universal normative claims without positing irreducible normative truths or grounding norms merely in subjective, variable elements of human psychology. Constitutivists therefore hope to sidestep a series of traditional objections to ethical theories. In addition, constitutivism provides a relatively straightforward explanation of why and how normative claims have their grip on us.

This essay offers an overview of the constitutivist strategy. Section One provides a constitutivist schema for extracting normative content from the notion of agency. Section Two sketches the versions of constitutivism that have been defended in the recent literature. Section Three looks more closely at the structure of constitutivism. Section Four examines the advantages that constitutivism has over more traditional ethical theories. Finally, Section Five considers a series of recent objections to constitutivism.

## 1. How constitutivism works

Constitutivists hope to show that action has a certain feature – a constitutive aim or principle – which is present in all episodes of action and engenders a standard of success for action. To see how this might work, it's helpful to begin with a simpler case: not action as such, but a particular *kind* of action. Certain kinds of action are distinguished by the fact that participants in these activities necessarily have certain aims. Games, such as chess, provide clear examples. Insofar as you play chess, you must aim at checkmating your opponent (or at least at attaining a draw). If you lack this aim—if you are simply moving pieces about on the board in accordance with the rules of chess, but are not aiming to checkmate your opponent—then you are not playing chess. It follows that the aim of checkmate is present in all episodes of chess playing. If you do not have this aim, you are not playing chess.

Let me offer an immediate qualification: checkmate is very rarely the *sole* aim of chess players. I may aim jointly at checkmating my opponent, having a good time, teaching my opponent a new move, engaging in a risky strategy, and so on. These aims may compete with and modulate one another: for example, I may see that I can attain checkmate in two moves, but decline to take these moves in order to prolong the game and try out a new strategy. Nonetheless, this kind of deviation from the aim of checkmate can only go so far, lest I cease to engage in the activity of chess playing. If I am not making *any* effort to attain checkmate, then I am not really playing chess.

An objection is likely to occur to many readers: surely it's possible to play chess with the intention of losing. After all, if I see someone playing chess with the intention of losing, I can correctly describe him as playing chess. So it seems that I can play chess without aiming at checkmate.

The constitutivist needn't deny this truism; she merely needs to point out that this is a case in which the agent is playing a *variant* of chess or a *more complex* version of chess, wherein the agent's additional aims modify or overrule the constitutive aim of checkmate. The important point is that describing the agent simply as *playing chess* in these cases elides important details. It's as if we describe the person interviewing for a job as *talking*—it's true, but it leaves out further aims that she has, aims that modify the activity in significant ways, and thereby require that someone who analyzes the features of the activity describe it more

precisely. Just so with chess: if you find out that your opponent is playing with the intention of losing, then in ordinary circumstances it does not matter whether we describe him as *playing chess* or *playing chess with the intention of losing*. However, the latter, more specific description picks out important features that distinguish this action from an ordinary case of playing chess. It's no surprise, and no objection, that the constitutivist will need to distinguish these cases. After all, the constitutivist isn't making a point about *how we typically describe* various actions; instead, she is analyzing the *structure* of these actions.

So the aim of checkmate seems to have a special status for chess players. Unlike the other, contingent aims that chess players might have (prolonging the game, enjoying the game, losing, etc.), this one cannot be set aside on pain of abandoning the activity in question. Let me clarify this point by introducing a bit of terminology. We can define *constitutive aim* 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.<sup>1</sup>

For example, suppose A is *chess playing* and G is *checkmate*. Then chess playing has a constitutive aim of checkmate iff (i) each token of chess playing aims at checkmate, and (ii) aiming at checkmate is part of what constitutes an event as an episode of chess playing. If the brief description above is correct, then these conditions seem to obtain, and checkmate is indeed the constitutive aim of chess playing.

What would follow from the fact that chess playing has a constitutive aim of checkmate? Well, suppose we accept a relatively minimal normative claim:

---

<sup>1</sup> Here it is worth making three points about the definition. First, condition (i) is in fact what does the work for constitutivists. If there were a way of establishing (i) without (ii), that would be fine. For, as we'll see below, it is the omnipresence of the aim that does the work; the constitution claim merely establishes the aim's omnipresence. Second, as these remarks imply, notice that condition (i) is implied by condition (ii). Strictly speaking, then, condition (i) is superfluous. I include (i) as a separate condition merely for the sake of clarity. Third, nothing important hinges on my restriction of A to attitudes and events; I would be happy to include other categories that might have constitutive norms. I cite attitudes and events simply because these are the categories that have been thought to possess constitutive norms. (I am including actions under the broader category of events.)

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

For example, if chess players aim at checkmate, then we can evaluate chess players with regard to whether their actions are conducive to their goal of achieving checkmate. If moving my rook would bring me closer to achieving checkmate, then I have reason to make that move. If the move would imperil checkmate, this provides me with a reason not to make it.<sup>2</sup>

So we have a very simple schema for generating normative claims from constitutive features. If an activity has a constitutive aim, then by appeal to Success we can generate normative conclusions. Of course, there are complications: we need to say why we should accept (some variant of) Success. And some constitutivists prefer to appeal to constitutive *principles* rather than constitutive *aims*. I will address these points in a moment.

For now, let's put these complications aside and consider how these points about chess might generalize. Constitutivists about action hope to show that action itself has a constitutive aim. If action had a constitutive aim, we could apply the Success principle to show that there is a standard of success pertaining to all actions, regardless of the particular goals that the agent pursues or the contingent motives that the agent has. The reasons generated by this standard would be *universal*: they would apply to all agents, regardless of the contingencies of the agent's beliefs, desires, and goals.<sup>3</sup> So, surprisingly enough, if action has a constitutive aim, and we accept some variant of Success, we can show that there are universal reasons.

## 2. Varieties of constitutivism

All well and good, but how are we to show that action has constitutive features? Chess is a game, defined by a rich set of rules and geared toward a particular outcome. It is

---

<sup>2</sup> For the moment, I am eliding a complication: we might think that the reasons generated by constitutive aims are *instrumental reasons*. For example, perhaps I have a reason to checkmate my opponent only if I have a reason to play chess. I address this complication below, in Section 5.4.

<sup>3</sup> They wouldn't necessarily be *categorical*, though. Categorical reasons apply independently of *any motive*. If we want to use constitutivism to establish categorical reasons, we need to avoid appeal to *aims*. Below, I discuss a version of constitutivism that does just this.

therefore *exactly* the right sort of activity to have a constitutive aim. But actions are decidedly heterogeneous: why think that walking to the store, writing a philosophy paper, deciding to take a nap, robbing a convenience store, chatting with a friend, flying to Chicago, going for a swim, brushing one's teeth, and listening to a concert all share some common aim? Why think they have *any* common property that is robust enough to generate interesting normative conclusions? This is the most challenging part of a constitutivist theory. The constitutivist must work with a plausible—or at least defensible—account of action and show that, according to this account, actions have a constitutive feature. In this section, I'll review the extant attempts to do so.

## 2.1 Velleman: action constitutively aims at self-understanding

David Velleman begins with an Anscombian conception of action, which we might state as follows:

1. If an agent intentionally A's, then the agent knows that she is A-ing and this knowledge is immediate.<sup>4</sup>

The knowledge is “immediate” in the sense that the agent does not need to introspect or observe evidence in order to determine what she is doing. When I intentionally walk to my office, I don't need to study my foot movements, introspect motives, and so on in order to determine what I am doing; I simply know, seemingly without relying on evidence, that I am walking to my office.

Suppose we accept this claim about intentional action. Velleman next argues for a particular theory of action that entails (1). In particular, he argues:

2. The best explanation for (1) is that agents have a desire to know what they are doing and fulfill this desire by forming self-fulfilling beliefs about what they will do.

---

<sup>4</sup> For his reliance on this conception of action, see for example Velleman (2009, 130-1).

This is initially counterintuitive and takes some explanation. Suppose an agent desires to know what she is doing. Then the agent will be averse to acting in ways that she wouldn't understand, and attracted to acting in ways that she would. For example, right now I know that I'd like a cup of coffee. Were I to walk to my kitchen and get a cup, I'd know what I was doing. Were I to walk to my kitchen and pour the pot of coffee down the sink, I wouldn't. So my desire to know what I am doing inclines me to perform the former action and avoid the latter.

As this example indicates, we can fulfill our desire to know what we are doing by modifying our behavior in light of our antecedent beliefs, rather than modifying our beliefs in light of our behavior. Velleman's core idea is that this mechanism is pervasive. He hypothesizes that, under the pressure of the desire to know what one is doing, agents form expectations about their forthcoming actions. Once an agent forms such an expectation, her desire to know what she is doing will incline her to act in a way that fulfills the expectation. For this reason, expectations will tend to be *self-fulfilling*: in an important range of cases, agents will be able to make it the case that they will A simply by forming the expectation that they will A.<sup>5</sup>

If intentional action were always caused by self-fulfilling expectations, then an agent would always possess a distinctive kind of knowledge of his own intentional actions: as Velleman puts it, the agent would attain "contemporaneous knowledge of his actions by attaining anticipatory knowledge of them" (2004b, 277; cf. 2009, 18-20). So Velleman asks us to entertain the hypothesis that intentional actions are just those behaviors that are caused by self-fulfilling expectations. This hypothesis recommends itself in two ways. First, it offers an explanation of the otherwise puzzling claim that intentional actions involve immediate knowledge (i.e., the knowledge in question is knowledge resulting from self-fulfilling beliefs about one's forthcoming actions). Second, it explains why actions always involve this distinctive kind of knowledge (i.e., intentional actions just are behaviors caused by self-fulfilling beliefs).

Interestingly, claim (2) entails that action has a constitutive aim. For:

---

<sup>5</sup> Velleman bolsters his case by surveying a range of empirical psychology that supports this claim. See Velleman (2006, Chapter Ten).

3. It follows that the etiology of every action includes a desire to know what one is doing, and that part of what it is for something to be an action is for it to have this desire in its etiology.
4. By (3), action has a constitutive aim of knowing what one is doing.

We can then derive normative conclusions by appeal to Success:

5. By (Success), agents have reason to act in a way that promotes knowledge of what they are doing.

In sum, then, Velleman argues that we can start with the idea that action is immediately known, show that this is best explained by the claim that action has a constitutive aim of knowing what one is doing, and generate normative conclusions from this claim. Put briefly: action constitutively aims at self-understanding (or, as Velleman sometimes puts it, at “intelligibility”).

## **2.2 Korsgaard: action’s constitutive principle is the Categorical Imperative**

Christine Korsgaard begins with the following conception of action:

1. An agent’s A-ing is an action iff A-ing is attributable to the agent as a unified whole.<sup>6</sup>

Just as Velleman started with a characterization of an essential feature of intentional action and then offered a theory of action to explain the presence of this feature, so too Korsgaard proceeds to offer a theory that explains why (1) is so. Her argument is complex, and I can only provide a brief outline here.

Korsgaard argues, on roughly Kantian grounds, that when self-conscious agents act, they experience a reflective distance from their motives (2009, 125-6). In order to decide how to act on the basis of competing motives, self-conscious agents must employ a “principle of choice” (2009, 119). For example, I have a desire to go for a walk and a desire

---

<sup>6</sup> See Korsgaard (2009, 18).

to continue working. I reflect on these desires and employ a principle such as “I’ll act on whichever desire is strongest” in order to determine what to do.

The principle that the agent employs determines the extent to which the agent she is unified (2009, 162-74). If I choose on the basis of a principle such as “I’ll act on my strongest desire,” this engenders one degree of unity; if I choose on a principle such as “I’ll do the honorable thing,” this engenders a different degree of unity. Korsgaard argues that the Categorical Imperative is the only principle that fully unifies the agent (2009, 78-81). With these claims in place, she can argue as follows:

2. An action is attributable to the agent as a unified whole iff the principle of the agent’s action unifies the agent.
3. The Categorical Imperative is the only principle that fully unifies the agent.
4. Therefore, an agent’s A-ing is an action iff the principle of the agent’s A-ing is the Categorical Imperative.
5. By (2)-(4), every action is governed by the Categorical Imperative.
6. By (2)-(5), being governed by the Categorical Imperative is part of what constitutes something as an action.
7. By (5) and (6), the Categorical Imperative is the constitutive principle of action.<sup>7</sup>

If this argument succeeds, it shows that every agent is committed to governing her actions in accordance with the Categorical Imperative.

### **2.3 Street: the constitutive features of the practical point of view**

Although Sharon Street describes her view as “constructivist,” it can be seen as a form of constitutivism that focuses on the constitutive features of *valuing* rather than *acting*. Street argues that the “truth of a normative claim consists in that claim’s following, in a logical or instrumental manner, from within the practical point of view, where the practical

---

<sup>7</sup> Notice that Korsgaard appeals to constitutive *principles* rather than constitutive *aims*. I explain this distinction in Section 3.2.

point of view is given a formal characterization” (2012, 40). So she aspires to provide an account of the practical point of view—the point of view of a deliberating agent. This account will be merely formal, in the sense that no substantive values are presupposed.

In particular, Street characterizes the practical point of view by focusing on the attitude of valuing, which she defines as “taking oneself to have a reason” (43). She argues that valuing has three key features. First, valuing obeys the instrumental principle: “*valuing* an end, in contrast to merely desiring it, constitutively involves *valuing* what one is fully aware is the necessary means to that end” (2012, 44). Second, valuing involves a wide range of emotions (2012, 44). Third, valuing has “greater structural complexity than the attitude of mere desiring.” For valuing involves “experiencing very specific features of the world as ‘calling for’ or ‘demanding’ or ‘counting in favor of’ other very specific things” (2012, 44).

With this account of valuing at hand, Street argues that the practical point of view is simply the point of view of an agent who values anything at all. Street accepts the Kantian claim that, as she puts it, “if you value something, then you cannot—simultaneously, in full, conscious awareness—also think that there is *no reason whatsoever* to value it” (2012, 46). So, for any particular valued end, we can ask what our reason for valuing it is. The Kantian aspires to show that this demand for justification leads to a regress that bottoms out in our commitment to the value of humanity. Yet Street rejects this move, claiming that we justify particular values simply by appealing to other values that we accept. We aspire to render these values consistent with one another, but there is no further ground to which we can appeal to justify the whole set of values (2012, 51). Accordingly, Street argues that we can seek nothing more than a “coherent web of interlocking values” (2012, 51). Coherence is the only standard to which we are committed by the constitutive features of the practical point of view.

#### **2.4 Smith: the constitutive features of fully rational agency**

Michael Smith takes a rather different approach. Rather than attempting to analyze the constitutive features of intentional action, Smith investigates the structure of a particular kind of intentional action: fully rational action. He suggests that there might be “certain desires [that] are constitutive of being fully rational” (2010, 134). For example, he tentatively

proposes that rationality requires that people intrinsically desire that there is as much happiness as possible (2010, 134).

How do we determine whether a desire is constitutive of rational agency? Smith argues that we simply engage “in a certain canonical method of reasoning: the process of reflective equilibrium” (2010, 136). In particular, “we can figure out which desires are constitutive of being rational by getting our considered judgments about what reason requires us to desire in specific cases...into equilibrium with our reflective judgments about what the most general intrinsic desires constitutive of being rational are that stand behind these more specific desires” (2010, 136-7). In short, reflective equilibrium reveals that certain desires are constitutive of rational agency. In work in progress, Smith pursues this strategy to argue that there are additional desires, with great moral import, that are constitutive of rational agency (see Smith [in progress]).

## **2.5 Katsafanas: action constitutively aims at agential activity**

I argue that action has two constitutive aims: (1) agential activity, which I analyze as requiring stable approval of one’s action given complete information about its etiology, and (2) encountering and overcoming resistance to one’s ends.

My account begins with a defense of the idea that there is a distinction between actions that spring from our own activity and actions that are brought about in a more passive fashion. Consider the difference between deliberately spurning someone and being led, by unrecognized feelings of anxiety, to avoid someone; arguably, the former case displays a kind of agential activity that is absent in the latter case. This is just one example, but there are many cases of this form.<sup>8</sup> Drawing on considerations about the nature of motivation and deliberation, I argue:

---

<sup>8</sup> Many action theorists agree that we need a distinction between two kinds of action. The distinction is described in different ways: it has been characterized by turns as identification, wholeheartedness, guidance by the agent, direction by the agent, agential control, agential activity, reflective self-control, rational control, and so forth. All of these terms have different connotations, and are analyzed in different ways. But many of the theorists who employ these terms agree on one point: these terms are meant to characterize the idea that agents can be more and less active in the production of their own actions. Accordingly, I use “agential activity” as the most general term for notions that are meant to pick out the agent’s contribution to the production of action.

1. An agent is active in the production of her action iff (a) the agent approves of her action and (b) further knowledge of the motives figuring in the etiology of this action would not undermine her approval of her action.

Employing a roughly Kantian argument about the nature of assuming our freedom from the practical standpoint, I argue that agents are committed to aiming at agential activity in every episode of reflective (or self-conscious) action, and part of what it is to for an action to be reflective is for it to have this aim. It follows that agential activity is the constitutive aim of reflective action.

On its own, this aim generates very little normative content. After all, it says nothing about the *basis* of the agent's approval of her action. However, I argue that if we investigate the structure of human motivation, we can uncover a second constitutive aim that provides more substance to the theory. Drawing on some insights from Nietzsche's analyses of human motivation, I argue that each episode of action aims not only at achieving some end, but also at encountering and overcoming obstacles or resistances in the pursuit of this end. For example, I aim not merely at writing this paper, but at writing it in a way that engenders difficulties to overcome. Although initially counterintuitive, I show that this account is supported by empirical and philosophical considerations about the nature of human motivation.

In particular, Nietzsche's account of agency is based on the idea that our actions are motivated by a distinctive kind of psychological state: the drive. Drives admit an aim/object distinction, where the aim is a characteristic form of activity and the object is a chance occasion for expression. For example, the sex drive aims at sexual activity, and might take as its temporary object a particular person. As the drive's objects are adventitious, there is there is no object the attainment of which completes or satisfies the drive. Rather than disposing us to acquire their objects, then, drives simply lead us to engage in their characteristic form of activity.

Elsewhere, I argue that this fact about drives entails that all drive-motivated actions share a common, higher-order aim of encountering and overcoming resistance. For part of what it is to be motivated by a drive—to manifest a drive's activity—is continuously to seek new objects on which to direct the drive's activity. In seeking new objects, we seek new resistances. In virtue of being motivated by drives, then, we seek continuous encountering

and overcoming of resistance, rather than the attainment of definite states. Nietzsche describes this fact about our aims by saying that we manifest “the will to power.”

This aim generates its own standard of success: we have reason to perform those actions that provide opportunities for encountering and overcoming resistance. And the constitutive aim of agential activity interacts with this constitutive aim in two ways. First, the standard generated by the aim of encountering and overcoming resistance gives us a way of assessing the values upon which activity depends: we reject those values that generate conflicts with this aim. Second, recall that agential activity requires both that the agent approve of her action and that this approval be stable in the face of further information about the action’s etiology. If the aim of encountering and overcoming resistance is indeed present in the etiology of every action, as the arguments above sought to establish, then an agent will be active only if the presence of this aim doesn’t undermine her approval of her actions. In Katsafanas (2013), I argue that a host of important normative conclusions can be generated by these claims.

### **3. The structure of constitutivism**

As the above summaries indicate, constitutivists begin by specifying some conception of action and then try to show that we can extract normative conclusions from it. Accordingly, constitutivist accounts can vary in three ways: the account of action with which they begin; their method for extracting normative content from this account of action; and the substantive conclusions about reasons that they allegedly generate. Below, I will treat these points in turn.

#### **3.1 The account of action**

As our survey of the extant versions of constitutivism indicates, constitutivists work with very different accounts of action. For example, whereas Velleman begins with the Anscombian idea that actions are immediately known, Korsgaard starts with the idea that that action is distinguished from mere behavior in that the former, but not the latter, issues from a unified agent. Accordingly, we could reject these accounts by objecting to the

accounts of action on which they are based. So, before constitutivism can even get off the ground, the constitutivist needs to show why we should accept the initial account of action.

In short: constitutivist theories can only be as plausible as the account of action upon which they are based. For this reason, the interest of a constitutivist theory is inversely proportional to the contentfulness of its initial conception of action. It's easiest to illustrate this with an example. Suppose we take some restricted practice, such as driving, and attempt to give a constitutivist account of the reasons pertaining to it. And suppose we say the constitutive aim of driving is *operating a vehicle while obeying the laws governing this activity*. These laws include: drive at the posted speed limit; stop at red lights; signal when changing lanes. It will be very easy to extract a rich array of reasons from this conception of driving: all drivers will have reason to obey the speed limit, stop at red lights, signal when changing lanes, and so on. But this is unsurprising, given that our initial conception of driving already builds in this normative content. If we started with a more minimal account of driving, such as *operating a vehicle*, then things would not be so straightforward.

And, of course, the more minimal account is in this case the more plausible one: it's obvious that you can drive without aiming to obey the laws governing this activity. So showing that we can extract normative content from the implausibly substantive account of driving is of no interest.

Just so with action. If we start with a very rich, substantive account of action, we will have an easy time extracting normative content from it. Yet this will be unsurprising and uninteresting. The most interesting version of constitutivism would start with an extremely minimal, uncontroversial account of action, and show that this account generates normative conclusions. For example, suppose we start merely with the idea that action is goal-directed movement, and show that we can somehow extract a claim such as "you have reason to help others" from this. This would be an astonishing and extremely powerful conclusion: it would move from a universally accepted account of action to a substantive normative conclusion. But suppose that, in order to extract the claim "you have reason to help others," we have to start with the idea that action is movement governed by the Categorical Imperative. This would be much less interesting, as all the work would be done in the defense of the initial account of action.

### **3.2 Extracting normative content from the account of action**

The second dimension on which constitutivist accounts vary is the way in which they extract normative content from their starting conceptions of action. One option is to show that action has a constitutive aim, and extract content from it by appealing to the idea that aims generate reasons. I've explained this strategy in Section One. But some constitutivists—especially those with Kantian sympathies—eschew this strategy.

*Kantian* versions of constitutivism view mere *aims* as inadequate for generating normative content; aims, along with associated motives, are (so the Kantian story goes) external to the will, so acting upon them would result in heteronomy.<sup>9</sup> Thus, when we turn to Korsgaard's Kantian version of constitutivism, we find reliance on a somewhat different constitutive feature: not a constitutive aim but a constitutive *principle* or *standard*. As the reconstruction of her argument in Section 2 indicates, Korsgaard makes no appeal to constitutive aims. Instead, she tries to show that action constitutively involves commitment to certain principles. If she succeeds in showing that each action is governed by the Categorical Imperative, then she can derive normative conclusions without appealing to a principle like Success.

So there are both aim-based and principle-based versions of constitutivism. Both seek to show that action has some constitutive feature that generates substantive normative content, but they go about this in different ways. It's helpful to think of these as Humean and Kantian variants of constitutivism, respectively. The Humean constitutivist shows that action has constitutive aims, and appeals to (something like) Success in order to derive substantive normative content. The Kantian constitutivist, by contrast, argues that action requires commitment to constitutive principles, and derives substantive normative content from the agent's commitment to these principles.

### 3.3 The substantive normative claims generated by the theories

We've seen that constitutivist theories differ both in the accounts of action and the manner in which they generate normative claims from these accounts. But there's a final

---

<sup>9</sup> Kant writes, "if the will seeks the law that is to determine it anywhere else than in the fitness of its maxims for its own giving of universal law...heteronomy always results" (*Groundwork* 4:441).

dimension to consider: constitutivist theories also differ in the content of the normative claims that they purport to generate.

A minimally ambitious version of constitutivism would claim that only the instrumental principle is constitutive of action. After all, we can derive a version of this principle from an elementary conception of action: to act is simply *to bring something about*. On this interpretation, the paradigmatic case of action has the following form: I desire some end X, I see that I could get X by doing Y, so I do Y. Action aims merely at effecting a change in the world, so that the world conforms to my desires. Some philosophers argue that we can move from this extremely minimal conception of action to a constitutivist account of the instrumental principle, which tells us to take the known and available necessary means to our ends.<sup>10</sup> The most minimal version of constitutivism would claim that the instrumental principle, and *only* the instrumental principle, can be given a constitutivist justification.

At the other end of the spectrum, a maximally ambitious version of constitutivism would claim that *all* of the moral claims that we typically accept can be given a constitutivist justification. Korsgaard pursues something like this strategy, claiming that the core of our current moral code can be extracted from the constitutive features of agency. As she puts it, we can derive “Enlightenment morality” from the constitutive features of action (1996, 123).

Of course, many constitutivist views fall somewhere between these two extremes. Take Velleman: although he denies that we can extract our moral code from facts about agency, he does think that we can give a constitutivist account of various norms that “favor morality without requiring or guaranteeing it” (Velleman 2009, 149). In short, we can get more than the instrumental principle but less than Enlightenment morality.

#### **4. The attractions of constitutivism**

We’ve reviewed the varieties of constitutivism and investigated the structure of constitutivist theories. Suppose some version of constitutivism works; suppose we can ground normative claims in facts about the constitutive features of action. Why would this matter?

---

<sup>10</sup> See, for example, Christine Korsgaard (1997), James Drier (1997), Michael Smith (2010), and Paul Katsafanas (2013, Chapter 1).

The approach would have several advantages over its competitors. First, it would provide a way of resolving the longstanding debate between internalists and externalists about practical reason.<sup>11</sup> Second, it would provide an easy answer to the question why normative claims should have any authority over us. Third, the theoretical commitments of the strategy are quite minimal. I'll address these points in turn.

#### **4.1 Constitutivism avoids the problems facing internalist and externalist theories of reasons**

Consider a perfectly homely normative claim, such as “you have to go to the movies.” If we ask what would render this claim true, the answer seems clear: a fact about the agent’s motives. If the claim is true for Allen but false for Betty, this is due to the fact that Allen desires to see the film and Betty does not. It is natural to think that in just this way, reasons will be tied to facts about agent’s motives.

But what about claims such as “you have reason not to murder”? That claim seems different. It purports to be *universal*, applying to all agents. Moreover, it does not seem to depend on the agent’s motives. Suppose Allen has many motives in favor of murdering his uncle (getting revenge for past slights, collecting an inheritance, etc.), and no motives that count against it (he’s a sociopath with no compunction about harming others, and he thinks he’s clever enough to contrive a plan that leaves him with no risk of getting caught). In this simplified case, all of Allen’s motives count in favor of murdering his uncle; none count against it. Nonetheless, most of us want to say that he has reason not to murder.

So we face contrary pressures: in certain cases, the claim that reasons are grounded in motives looks exceedingly plausible, indeed obvious; in others, the same claim looks like it generates unacceptable consequences. And so we get a familiar, well-worn philosophical debate: *internalists* defend the claim that all normative claims are generated in facts about the agent’s motives, whereas *externalists* deny this. More precisely:

(Internalism) Agent A has reason to  $\phi$  iff A has, or would have after procedurally rational deliberation, a desire or aim whose fulfillment would be promoted by  $\phi$ -ing.

---

<sup>11</sup> I discuss this topic in Katsafanas (2011), and some of the material in Section 4.1 is drawn from there.

(Externalism) It can be true both that (i) agent A has reason to  $\phi$ , and (ii) A does not have, and would not have after procedurally rational deliberation, a desire or aim whose fulfillment would be promoted by  $\phi$ -ing.

Each of these theories faces certain difficulties. Internalism has trouble with apparently universal normative claims, such as “you should not murder.”<sup>12</sup> Externalism is tailor-made to capture universal normative claims. Nonetheless, it faces several challenges, including the much-discussed problems of *practicality* and *queerness*. First, consider practicality. Moral claims are supposed to be capable of *moving us*. Recognizing that  $\phi$ -ing is wrong is supposed to be capable of motivating the agent not to  $\phi$ . But we might wonder how a claim that bears no relation to any of our motives could have this motivational grip. As Bernard Williams puts it, “the whole point of external reasons statements is that they can be true independently of an agent’s motivations. But nothing can explain an agent’s (intentional) actions except something that motivates him so to act” (1981, 107). William’s suggestion is that if the fact that murder is wrong is to exert a motivational influence upon the person’s action, then the agent must have some motive that is suitably connected to not murdering. And this pushes us back in the direction of internalism.

Second, consider Mackie’s argument from queerness. Motives are familiar things, so it seems easy enough to imagine that claims about reasons are claims about relations between actions and motives. Internalism therefore has little difficulty with Mackie’s argument. But what would the relation in an external reasons statement be? Are we to imagine that a claim about reasons is a claim about a relation between an action and some independently existing value? This would be odd: as Mackie puts it, “if there were objective values then they would be entities or relations of a very strange sort, utterly different than anything else in the universe” (1977, 38). For if such values existed, then it would be possible for a certain state of affairs to have “a demand for such-and-such an action somehow built into it” (1977, 40). And this, Mackie concludes, would be a decidedly queer property.

In sum: both externalism and internalism have attractive features, yet incur substantial costs. Traditional internalism grounds normative claims in familiar features of

---

<sup>12</sup> Of course, internalists have attempted to modify the view in order to avoid this problem. For two classic examples, see Smith 1994 and McDowell 2001.

our psychologies, yet for that very reason has trouble generating universal normative claims. Externalism generates universal normative claims with ease, yet encounters the problems of practicality and queerness. So we have a pair of unappealing options, and the debate continues.

Constitutivism attempts to resolve this dilemma. To put it in an old-fashioned way, constitutivism sublates internalism and externalism, seeing each position as containing a grain of truth, but also as partial and one-sided. The constitutivist agrees with the internalist that the truth of a normative claim depends on the agent's aims, in the sense that the agent must possess a certain aim in order for the normative claim to be true. However, the constitutivist traces the authority of norms to an aim that has a special status—an aim that is *constitutive* of being an agent. This constitutive aim is not optional; if you lack the aim, you are not an agent at all. So, while the constitutivist agrees with the internalist that reasons derive from the agent's aims, the constitutivist holds that there is at least one aim that is intrinsic to being an agent. Accordingly, the constitutivist gets one of the conclusions that the externalist wanted: there are universal reasons for acting.<sup>13</sup> Put differently, there are reasons for action that arise merely from the fact that one is an agent. Specifically, these are the reasons grounded in the constitutive aim.

So constitutivism can be viewed as an attempt to resolve the dispute between externalists and internalists about practical reason, by showing that there are reasons that arise from non-optional aims.<sup>14</sup> In so doing, it generates universal reasons while sidestepping the problems of practicality and queerness.

## 4.2 The authority problem

Normative claims make demands on us: they tell us which actions to perform and which to avoid. But what justifies this authority? Put differently, if a skeptic asks why he should accept of the form “you have reason to  $\phi$ ,” what can we say to him? Constitutivism has an easy way of answering this question: insofar as you are committed to an activity with

---

<sup>13</sup> Which is not to say that the constitutivist gets *everything* that the externalist might want. For example, as I explain in the final section, the constitutivist avoids a realist metaethics, and some externalists will find this unsatisfying.

<sup>14</sup> For this reason, constitutivism has been called both “meta-internalism” and “quasi-externalism” (Wallace 2006, Chapter Three; Enoch 2006, 172).

constitutive standards, you are committed to those standards. For example, a person who is playing chess has a good reason to govern her activities with the standard of achieving checkmate: if she doesn't, she won't be playing chess.

To see why this is important, it helps to contrast standards generated by constitutive aims with other types of standards. Consider a normative claim such as “you should obey the posted speed limit.” Suppose someone asks why he should conform to this standard. Of course, there are answers to this question—we can invoke the financial penalties that the exceeding the speed limit may produce, the danger to the other drivers created by excessive speeds, respect for the law, and so on. But notice that these answers invoke *external* standards. The standards apply because obeying the speed limit is related to other activities, goals, and practices that the agent accepts or cares about. The standard of checkmate doesn't need to rely on these external features: we can answer the question “why should I care about capturing your queen?” simply by referring to the activity's constitutive aim. Thus, the chess player should care about capturing the queen because if she doesn't govern herself by the standard of achieving checkmate, she won't be playing chess. So this is the second intriguing feature of constitutive standards: they are internal to the activities that they regulate. Accordingly, we need not invoke external facts in order to legitimate their claim to authority.<sup>15</sup>

### 4.3 The minimal theoretical commitments of constitutivism

Third, notice that constitutivism makes very minimal assumptions about normativity. In generating claims about reasons, constitutivists do have to rely on some claim linking normativity to the agent's aims or principles; Success, defined above, is one option, and versions of it are adopted by many constitutivists (Velleman 2000; Street 2012; Smith 2010; Katsafanas 2013). While not entirely uncontroversial, Success is an exceedingly spare claim. It can serve as an area of overlapping consensus in debates about ethics: we disagree about whether we have reason to be compassionate, whether happiness is more important than duty, whether suicide is wrong, and so forth. But we can set aside this disagreement on

---

<sup>15</sup> External facts will be relevant, of course. If I am engaged in a game of chess, and suddenly notice that my house is burning down around me, then there's a very real sense in which my reasons for moving my rook are outweighed or silenced by my reason to stop playing and call the fire department.

substantive ends and agree on this entirely procedural or structural conception of rationality: we can agree that if you have an aim, you should strive to fulfill it, while disagreeing about what those aims are. Or, if we are considering Korsgaard's principle-based version of constitutivism, we can say that if you participate in a practice with a constitutive principle, you should strive to conform to that principle. (I'll consider an objection to this claim in the next section.)

## **5. Some problems for constitutivist theories**

We have seen that constitutivism is attractive because it sidesteps metaphysical and practical problems, generates claims about reasons without invoking goals or standards external to the activity in question, and yields normative claims while relying only on the most minimal of normative assumptions. Constitutivism thus has considerable promise. However, it also faces a number of difficulties. Some of these are internal to particular versions of the theory; others affect constitutivism as such. Below, I will offer a brief review of seven central problems for constitutivist theories.

### **5.1 Equivocation on the concept of action**

Let's start with an internal difficulty. I mentioned above that the more content we build into our conception of action, the easier it will be to extract normative content from its constitutive features, but the less interesting this project will be. This gives rise to a danger: it's perennially tempting for the constitutivist to start with a very minimal account of action but, over the course of the argument, to surreptitiously switch to a richer conception of action.

After all, there are two stages to a constitutivist theory. First, the constitutivist defends some account of action. Second, the constitutivist argues that we can extract normative content from this account of action. The first part of the constitutivist argument is easiest if we work with minimal accounts of action; the second part is easiest if we work with rich, substantive accounts of action. So when arguing for their conceptions of action, constitutivists will be under pressure to work with very minimal accounts of action; when engaged in extracting claims about reasons, they'll be under pressure to work with much

richer accounts of action. And the danger is that, in the face of these pressures, the constitutivists will equivocate, relying on one conception of action in the first stage and another at the second stage.

This is no idle worry: I've argued that Korsgaard and Velleman succumb to exactly this pressure (Katsafanas 2013, Chapters 3-4; see also Tiffany 2011). The problem is clearest in Velleman's theory. He begins with the claim that action is immediately known. This isn't completely uncontroversial, but it's widely accepted and we can give good arguments for it. And he points out that we can account for the presence of immediate knowledge by postulating that whenever an agent acts, the agent has a desire to know what she is doing; this desire, he points out, would lead the agent to act in a way that she antecedently expects to act. This seems fine: the account of action may look counterintuitive initially, but Velleman gives good arguments for it and the basic claim turns out to be supported by empirical psychology. But I've argued that there is a problem: Velleman wants to show not just that we have reason to act in a way that generates self-knowledge, but that we have *more* reason to act in ways that generate *more* self-knowledge. How do we get there? From the fact that each action is motivated by a desire for *some* self-knowledge, it does not follow that we have *more* reason to act in ways that generate *more* self-knowledge. To get that conclusion, we'd have to show that each action is motivated by a desire to know *as much as possible* about what one is doing. In other words, we need to move from a comparatively minimal claim that agents aim to have some knowledge of what they're doing to the very strong claim that agents aim to have comprehensive knowledge of what they're doing. And it is not clear that Velleman's arguments support this stronger conception of action; I've argued that they support only the weaker conception. (I argue that an analogous problem plagues Korsgaard's theory, though given the complexity of her argument I lack the space to reconstruct the problem here. See Katsafanas [2013, Chapter Four] for the details; see also Tubert [2011].)

## 5.2 Agency, schmagency

Constitutivists aspire to show that we can generate universal reasons from an account of action. But reflect for a moment on the constitutivist account of chess discussed above. Suppose I see that two agents appear to be playing chess but are violating one of the

rules: they're moving their rooks on diagonals. I point out to them that they're playing a defective game of chess. They respond by saying that they've decided to modify one of the rules of chess: you can move your rooks on diagonals.

A stickler for constitutivist characterizations of practices, I tell them that they're not really playing *chess* at all: they're playing a slight variant of chess. Call it *schmess*. Given that they're not playing chess, the reasons generated by the constitutive features of chess don't apply to them. For example, a chess player would be violating a norm if she moved her rook on a diagonal. But not a *schmess* player; that move is fine.

I use this example to illustrate a more general point. For any activity with a constitutive feature, you can escape the reasons generated by that constitutive feature by engaging in a slight variant of the activity that doesn't include the relevant feature. David Enoch has argued that the same point applies to action as such. Suppose action has a constitutive standard of self-constitution (as Korsgaard claims). If I don't care for this constitutive standard, perhaps I can simply engage in a slight variant of action; call it *schmaction*. *Schmaction* is just like action, except that it doesn't include the constitutive standard of self-constitution.

Enoch argues that this possibility shows that constitutivism fails. For, presented with the idea that, say, self-constitution is constitutive of agency, someone can respond, "I am perfectly happy performing *schmactions*—nonaction events that are very similar to actions but that lack the aim (constitutive of actions but not of *schmactions*) of self-constitution" (Enoch 2006, 179). If that is right, then the reasons generated by constitutive aims are not universal after all; they apply only to those who decide to engage in particular forms of action.

Does this objection succeed? It depends. Enoch's objection rests on the idea that it is possible to perform *schmactions* instead of actions. But this is just the possibility that constitutivists want to deny. After all, the constitutivist aspires to begin with an account of intentional action as such. That is, by "action" the constitutivist just means *intentional activity*. Any intentional activity that the agent performs will count as an action. Thus, the idea that there could be a *schmaction*—an intentional activity that is not an action—is contradictory.

But Enoch is on to something. I pointed out above that it's perennially tempting for the constitutivist to begin with a very minimal account of action, but switch to a more substantive one as the argument progresses. If the constitutivist does so—if the

constitutivist isn't really relying on a conception of mere intentional activity, but instead of movement governed by the categorical imperative (or some such)—then it seems that an agent *can* decline to engage in action-so-described while still engaging in intentional activity. In other words, if the constitutive features are present only in some more refined form of agency, then I can escape them by engaging in a less refined form of agency.

To put the point briefly: if action, as the constitutivist analyzes it, is inescapable, then the schmagency objection fails. If action is escapable—if the agent can engage in intentional activity without meeting the constitutivist's criteria for acting—then the schmagency objection succeeds. So the lesson of the objection is that the constitutivist needs to show that action, as she analyzes that notion, is inescapable.

### 5.3 Inescapability and normativity

I've just argued that if the constitutivist can establish that action, as she analyzes it, is inescapable, then she can avoid the schmagency objection. But Enoch has recently attacked this kind of response as well. He writes,

The move from “You inescapably  $\Phi$ ” to “You should  $\Phi$ ” is no better—not even that tiniest little bit—than the move from “You actually  $\Phi$ ” to “You should  $\Phi$ ”.  
(Enoch 2011, 216)

For example, suppose the alcoholic has an inescapable craving for more alcohol. It hardly seems to follow that he should take another drink. So inescapable tendencies don't always generate reasons. Does this imply that the constitutivist is making a grave error?

Not at all. The first thing to notice is that Enoch has misstated the constitutivist strategy. The constitutivist doesn't move from “you inescapably  $\Phi$ ” to “you should  $\Phi$ .” Rather, the aim-based version of constitutivism moves from “you inescapably *aim at*  $\Phi$ -ing” to “you should  $\Phi$ .”<sup>16</sup>

This is a crucial distinction. The mere fact that I *do* something doesn't generate standards of success. If I catch my foot on a crack in the sidewalk and stumble, it does not

---

<sup>16</sup> Notice that the move is not from “you inescapably aim at  $\Phi$ -ing” to “you should *aim at*  $\Phi$ -ing”. Rather, it is from “you inescapably aim at  $\Phi$ -ing” to “you should  $\Phi$ .” Analogously, the principle-based version of constitutivism moves from “you inescapably engage in an activity governed by principle P” to “you should perform actions that fulfill P.”

follow that there is any standard of success for my stumbling. If I forget where I placed my keys, it does not follow that there is a standard of success for forgetting where my keys are.

However, Success claims that *aims* generate standards of success. If I *aim* to catch my foot on a crack in such a way that it causes me to stumble (perhaps I want to feign an injury, or perhaps I'm acting in a play) then there *is* a standard of success. And if the alcoholic aims to take another drink, he does have a reason to take one—a reason that, to be sure, may be outweighed by his other aims. So Enoch's objection misses its mark. (We might construe Enoch as objecting to the Success claim itself; I consider this possibility in the next section.)

Moreover, notice that there are different ways of formulating the Success claim. We might say that it's the *inescapability* that generates the normativity: inescapable aims, and only inescapable aims, generate standards of success. Or we might adopt a weaker claim, saying that all aims generate standards of success. Constitutivists only need the weaker claim. Constitutive aims differ from ordinary aims only in that constitutive aims are inescapable, whereas ordinary aims are not. It isn't the inescapability that is reason-providing. The aim itself—any aim—is reason-providing. The inescapability vouchsafes the reasons' universality, not their status as reasons.

#### 5.4 Aims and reasons

However, there is a potential objection: what if we reject the very idea that aims generate reasons? Consider reprehensible activities that have constitutive standards, such as terrorism.<sup>17</sup> Let's stipulate that terrorism aims at achieving a political goal by inflicting mass casualties on a civilian population. A standard of success for terrorist activity, then, is that it inflicts mass casualties. This means the terrorist has reason to (for example) construct a bomb, plant it in a crowded place, and detonate it. But do we really want to say that the terrorist has *reason* to do these things? Some philosophers are uneasy about this kind of conclusion. Thus, in a number of debates about instrumental reasons, some philosophers have defended the idea that one's end generates reasons to take the means only if one has reason to have the end.<sup>18</sup> For example, the terrorist would have reason to kill civilians only

---

<sup>17</sup> Cohen (1996) raises an objection of this form for an earlier version of Korsgaard's theory.

<sup>18</sup> For this version of the instrumental principle, see for example Raz (2005). He states his "facilitative principle" as follows: "When we have an undefeated reason to take an action, we have reason to perform any

if he antecedently had reason to be a terrorist. Others have argued that the instrumental principle expresses a rational requirement of the following form: rationality requires that if you aim at end E, then [either you give up this end or you take the necessary and available means to E]<sup>19</sup>. This would give us space to argue that the terrorist should drop his end of terrorism rather than kill civilians.

In other words, I have presented the constitutivist as relying on some version of the following claim:

(A) If you are  $\phi$ -ing, and  $\phi$ -ing constitutively aims at E, then you have reason to E.

But drawing on disputes about practical reason, we might reject (A) and argue for one of the following claims:

(B) If [you are  $\phi$ -ing, and you have reason to  $\phi$ , and  $\phi$ -ing constitutively aims at E], then you have reason to E.

(C) Rationality requires that if [you are  $\phi$ -ing, and  $\phi$ -ing constitutively aims at E], then [either you stop  $\phi$ -ing, or you take the necessary and available means to E]

These issues are hotly disputed, so it would be nice if the constitutivist could avoid taking a stand. So let's consider whether she can.

Claim (C) does not seem to present any problems for the constitutivist: after all, no one will deny that you can avoid the constitutive aim of chess by ceasing to play chess. And when we apply this point to action itself, it would follow that we are rationally required to take the necessary and available means to fulfilling the constitutive aim. Although the constitutivist project would be recast as involving rational requirements rather than reasons,

---

one (but only one) of the possible (for us) alternative plans that facilitate its performance" (2005, 5-6; italics removed).

<sup>19</sup> For this version of the instrumental principle, see for example Broome (1999). Broome calls this a wide-scope requirement: it expresses a requirement not to have certain combinations of attitudes, rather than to perform particular actions. **See Kolodny (this volume) for a discussion of these matters.**

and would then need to address the connection between the two, there is no obvious reason for thinking that this would be problematic.

Now consider (B). This seems more problematic for the constitutivist. After all, if the constitutivist needs to show that we have a reason for action as such, then the normative conclusions generated by the constitutive aim of action will be contingent: they will be dependent on one's having a reason to act. This would undermine the constitutivist's attempt to show that we can generate universal reasons merely from the notion of action. However, it is open to the constitutivist to argue that even if the reason-providing force of an activity's constitutive aim were *in general* dependent upon one's having a reason to engage in the activity, this could not be true when the activity in question is action itself. For action is crucially different from other, more particular types of activities with constitutive aims. In particular, action is unavoidable. This renders the question of whether there is a reason for action as such very odd: what exactly is the agent supposed to be asking when she poses this question? If the question is either incoherent or unstable, the constitutivist may be able to deny (C) in the case of action itself. (For arguments of this form, see Ferrero [2009], Velleman [2009], and Katsafanas [2013]).

## 5.5 The bad action problem

The constitutivist wants to move from the claim

- (i) action has constitutive feature F

to the claim

- (ii) F is the standard of success for action.

Railton (1997) and Clark (2001) have argued that this creates a problem: according to (i), every action has F; according to (ii), F is the standard of success for action. It follows that every action is a success. In other words, it seems that there is no distance between something's being an action and its being a good action.

The constitutivist can respond to this problem by drawing a distinction. Aiming at F (or being governed by standard S) is what constitutes an event as an action. But achieving F (or being in conformity with S) is what constitutes an event as a *good* action. Compare chess: you don't have to *achieve* checkmate to count as playing chess. You just have to *try*.

In principle, then, the constitutivist can account for bad action. But she will need an account of constitutive features that allows them to be differentially realizable. Velleman aspires to fulfill this demand by arguing that self-understanding comes in degrees; just so, Korsgaard argues that agential unity comes in degrees.

## 5.6 The maximization problem

In order to leave room for bad action, we need to introduce a differentially realizable constitutive feature. However, this generates a new problem. Suppose that aiming at self-understanding is constitutive of action. Then every episode of action will aim at self-understanding. However, it certainly seems possible to aim at self-understanding while regretting or disavowing the fact that I do so: perhaps I'd rather be ignorant of certain aspects of myself. It also seems possible to aim at self-understanding, but to care very little about the degree to which I fulfill this aim: self-understanding just isn't very important to me, and I'll be satisfied with the merest scrap. After all, I can aim at knowing about quantum physics or the second World War, but be satisfied with amateurish, popularized accounts of each. Just so, it seems that I can aim at knowing about myself, without bothering to acquire a very deep or comprehensive knowledge of myself.

If all that it takes to count as an action is to *aim* at self-understanding—and maybe even achieve a *bit* of it—then what commits the agent to achieving substantial self-understanding? This question is important for the constitutivist, because good actions are supposed to be those actions that fulfill the constitutive aim (or meet the constitutive principle) *completely*. As Velleman puts it, “reasons will have to qualify as better or stronger in relation to the constitutive aim of action, which lends reasons their normative force. Roughly speaking, the better reason will be the one that provides the better rationale—the better potential grasp of what we are doing” (Velleman 2000, 29). And Korsgaard writes that defective action is possible because “an action can unify and constitute its agent to a greater or lesser degree” (Korsgaard 2009, 163). Thus, Korsgaard claims that what is required for

something to be an action is that it constitute the agent as unified to *some* extent. But what is required for something to be a *good* action is that it constitute the agent as *fully* unified.

If our commitment to acting simply generates a commitment to realizing the constitutive standards *to some degree*, but good action is action that fulfills the aim *either completely or to a high degree*, then it seems that our commitment to action is not yet a commitment to *good action*. In order to avoid this problem, the constitutivist will have to show that our commitment to action somehow generates a commitment to realizing the constitutive features to a high degree. This, I think, is a severe problem for the constitutivist. Indeed, I have elsewhere argued that it is at this stage that Korsgaard's and Velleman's theories begin to equivocate on their conceptions of action, as I explained above in Section 5.1.

## 5.7 Constitutivism and traditional metaethics

A final problem concerns the aspirations of certain constitutivist theories. Some constitutivists claim that their theories provide an alternative to the familiar metaethical categories. Korsgaard, for example, insists that her theory renders traditional metaethics “boring” (Korsgaard 2008, 325 fn. 49). Velleman writes that he has “doubts about the way metaethics is generally practiced” (2009, 157). He sees his constitutivist approach as offering an alternative to extant metaethical theories, an alternative which shows that normative claims are “objectively valid by being inescapable, in the sense that any agent must accept them” (2009, 116).

So constitutivism is sometimes presented either as an alternative to metaethics or as a distinctive kind of metaethic. However, this claim has come under criticism by Silverstein (2012), Hussain (2012), and Hussain and Shah (2006). Constitutivism is broadly anti-realist, in the sense that it views normative claims as constituted by the features of agency. For example, a constitutivist account of chess doesn't claim that there are objective, practice-independent reasons to checkmate your opponents that are somehow tracked by the constitutivist account of chess; rather, it's the fact that chess players constitutively aim at checkmate that makes claims such as “you have reason to capture the queen” true. In this sense, constitutivism avoids commitment to a realist metaethic.

Perhaps this is all that Korsgaard and others mean when they claim that constitutivism avoids metaethics. However, this is a merely negative conclusion: it tells us what constitutivism *isn't*, not what it is. And once we try to say a bit more about what constitutivism is, it seems to line up rather nicely with familiar metaethical categories. For example, Silverstein (2012) forcefully argues that constitutivism is committed either to reductionism or to expressivism. Expressivism claims that the meanings of normative claims should be understood in terms of the mental states they express, and that these mental states are conative rather than cognitive (see Blackburn 1998 and Gibbard 2003). Reductionism of the relevant sort claims that the meaning of normative claims should be understood in terms of the commitments that we inescapably encounter as agents. If Silverstein and others are correct, then constitutivists will inherit the problems facing expressivist or reductionist metaethics.

## **6. Conclusion**

Constitutivism aspires to ground universal normative claims in facts about the nature of action. In this essay, I've outlined the constitutivist strategy, sketched the arguments for the extant versions of constitutivism, considered the advantages of the constitutivist approach, and reviewed a series of objections to the theory. I've suggested that while particular constitutivist theories might succumb to some of these objections, constitutivism as such retains its promise.

## References

- Berteau, Stefano. (2013), "Constitutivism and Normativity: A Qualified Defence," *Philosophical Explorations* 16(1): 81-95.
- Blackburn, Simon. (2001), *Ruling Passions*. Oxford: Oxford University Press.
- Broome, John. (1999), "Normative Requirements," *Ratio* 12(4): 398-419.
- Clark, Philip. (2001), "Velleman's Autonomism," *Ethics* 111: 580-593.
- Cohen, G.A. (1996), "Reason, Humanity, and the Moral Law," in Christine M. Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press), 167-188.
- Dreier, James. (1997), "Humean Doubts about the Practical Justification of Morality," in Cullity and Gaut (eds.), *Ethics and Practical Reason*. Oxford: Oxford University Press.
- Enoch, David. (2006), "Agency, Schmagency: Why Normativity Won't Come from What is Constitutive of Action," *Philosophical Review* 115: 169-198.
- (2011), "Schmagency Revisited," in Michael Brady (ed.), *New Waves in Metaethics* (New York: Palgrave Mcmillan), 208-233.
- Ferrero, Luca. (2009), "Constitutivism and the Inescapability of Agency," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics IV*: 303-333
- Gibbard, Alan. (2003), *Thinking How to Live*. Cambridge: Harvard University Press.
- Hussain, Nadeem J.Z. (2012), "A Problem for Ambitious Metanormative Constructivism," in J. Lenman and Y. Shemmer (eds.), *Constructivism in Practical Philosophy*. Oxford: Oxford University Press.
- Hussain, Nadeem J.Z. and Shah, Nishi. (2006), "Misunderstanding Metaethics: Korsgaard's Rejection of Realism," in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics Volume 1*. Oxford: Oxford University Press.
- Katsafanas, Paul. (2011), "Deriving Ethics from Action: A Nietzschean Version of Constitutivism," *Philosophy and Phenomenological Research* 83: 620-60.
- (2013), *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press.
- Korsgaard, Christine M. (1996), *The Sources of Normativity*. Cambridge: Cambridge University Press.
- (1999), "Self-Constitution in the Ethics of Plato and Kant," *Journal of Ethics* 3: 1-29.
- (2008), *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.
- (2009), *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- McDowell, John. (2001), *Mind, Value, and Reality*. Cambridge: Harvard University Press.
- Railton, Peter. (1997), "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action," in Cullity and Gaut (1997)
- Raz, Joseph. (2005), "The Myth of Instrumental Rationality," *Journal of Ethics and Social Philosophy* 1: 1-28.
- Silverstein, Matthew. (2012), "Inescapability and Normativity," *Journal of Ethics and Social Philosophy* 6(3): 1-26.
- Smith, Michael. (1994), *The Moral Problem*. Oxford: Wiley-Blackwell.
- (2010), "Beyond the Error Theory," in *A World Without Values: Essays on John Mackie's Moral Error Theory*, edited by Richard Joyce and Simon Kirchin (New York: Springer), 119-139.
- (in progress) "A Constitutivist Theory of Reasons"

- Street, Sharon. (2008), "Constructivism about Reasons," *Oxford Studies in Metaethics Volume 3*, Russ Shafer-Landau (ed.). Oxford: Oxford University Press.
- (2010), "What is Constructivism in Ethics and Metaethics?," *Philosophy Compass* 5(5): 363-84.
- (2012), "Coming to Terms with Contingency: Humean Constructivism about Practical Reason," in James Lenman and Yonatan Shemmer (eds.), *Constructivism in Practical Philosophy*. Oxford: Oxford University Press.
- Tiffany, Evan. (2011), "Why be an Agent?," *Australasian Journal of Philosophy* 90(2): 223-33.
- Tubert, Ariela. (2010), "Constitutive Arguments," *Philosophy Compass* 5: 656-666.
- (2011), "Korsgaard's Constitutive Arguments and the Principles of Practical Reason," *Philosophical Quarterly* 61: 343-62.
- Velleman, J. David. (2000), *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- (2009), *How We Get Along*. New York: Cambridge University Press.