4-13-2011

Autonomy, de facto and de jure

Paul Tulipana
Georgia State University

AUTONOMY, DE FACTO AND DE JURE

by

PAUL TULIPANA

Under the Direction of Christie Hartley and Sebastian Rand

ABSTRACT

On a standard philosophical conception, being autonomous is roughly equivalent to having some particular natural capacity. This paper provides argues that this conception is incorrect, or at least incomplete. The first chapter suggests that adopting an alternative conception of autonomy promises to resolve to several objections to the metaethical constitutivism, and so promises to provide highly desirable theory of moral reasons. The second chapter first motivates a broadly Kantian account of autonomous action, and then gives reasons to think that Kant's own development of this theory runs into damaging action-theoretic problems. The way to address these problems, I argue, is to modify Kant's account of autonomy in a way that leaves no room for the standard conception of autonomy to do any work.

INDEX WORDS: Autonomy, Agency, Action, Freedom, Constitutivism, Metaethics, Normativity, Apriority, Immanuel Kant, GWF Hegel

AUTONOMY, DE FACTO AND DE JURE

by

PAUL TULIPANA

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2011

AUTONOMY, DE FACTO AND DE JURE

by

PAUL TULIPANA

Committee Chairs: Christie Hartley

Sebastian Rand

Committee: Andrew Altman

Eddy Nahmias

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2011

ACKNOWLEDGEMENTS

Thanks to Andrew Altman, Justin Coates, Eddy Nahmias, Cindy Phillips, Hunter Thomsen, and the participants at the Fall 2010 Georgia Philosophical Society meeting for their helpful comments on this paper and on earlier versions of these arguments. Thanks especially to Christie Hartley, Sebastian Rand, and Eric Wilson, who got the worst of it. For their patience and insight, I am very grateful.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION : AUTONOMY, *DE FACTO* AND *DE JURE*

What does it mean to say that Jones acts autonomously? The most common philosophical answers to this question have something like the following form: it means that in acting, Jones exercises a certain natural capacity for, disposition to, or inclination toward self-governance, toward (self-)conscious control of her behavior, or toward determining the content of her activity in accordance with the deliverances of her practical reasoning, and that her acting is the result of, or is identical to, her doing that.[1] What, then, on this account, does it mean to say that Jones is autonomous? It means that Jones possesses the requisite capacity, disposition, or inclination.

The general goal of this paper is to show that there is something wrong, or at least importantly incomplete, about this conception of autonomy. More specifically, I will argue that this conception is problematic because it involves conceiving the autonomy-making feature of agents non-normatively.[2] Surely, some set or other of autonomy-enabling natural features is necessary for autonomy, but such inclinations, dispositions, capacities, and so on – whatever they turn out to be – are jointly insufficient for any being's successfully achieving the status of autonomy; so they cannot be the autonomy-making features of autonomous beings.

There are a variety of considerations that speak in favor of this claim. First, no naturalistic account of agency-making properties is very widely accepted. Although this is hardly conclusive that there is a problem with this way of thinking generally, it should recall the order of explanation operative in this area of inquiry. We do not think of and treat each other as autonomous

[1] There have been many specific proposals about the natural property that serves as the autonomy-maker for autonomous agents in the philosophical literature on agency. Just to name a few: a basal desire for self-understanding; an inclination toward autonomy itself; a disposition toward coherence; a particular set of "moral powers" and corresponding highest-order interests; and certain standing second-order desires. See, respectively, Velleman 2009, Velleman 1996, Smith 1987, Rawls 1980, and Brandt 1979. This note recapitulates a helpful review of "autonomy-making" capacities and motives in Rosati 2003, pp. 512ff.

[2] Most often, naturalistically.

because we have discovered that we are as a matter of natural fact possessed of certain capacities, dispositions, and so on. Rather, in attempting to characterize some of these latter as autonomy-making, we are casting about for a way of analyzing and justifying our already-in-place behavior. The idea that the autonomy-making feature is a fundamentally natural one, as opposed to a fundamentally normative one, is introduced at an early stage, but there is no compelling reason to think it is necessary for an analysis of autonomy to come off successfully. So there is at least some prima facie reason to think that we ought not simply *assume* naturalism about autonomy. Except perhaps that a certain kind of naturalism is historically coordinated with a certain conception of freedom.

Second, and more importantly, thinking of autonomy as a normative status has significant metaethical, ethical, and action-theoretic benefits. The first chapter of this paper will focus on a specific metaethical virtue of this conception of autonomy; the second will focus on a cluster of related action-theoretic virtues. On an exegetical register, these chapters can also be understood as the negative and positive aspects (respectively) of my reading of a kind of practical philosophy broadly associated with German Idealism. To speak anachronistically, the first chapter is a critique of Kant's metaethical constitutivism, and the second is an appreciation of Hegel's development of Kant's theory of autonomous action.

Before I start, I'll say something brief about the alternative to the conception of autonomy I have laid out as the standard one. The alternative I have in mind can be understood by appeal to a distinction about the appropriateness of predicates. Some predicates, I suggest, are appropriately applied de facto and some are appropriately applied de jure. Of course, this is itself a normative distinction. In some sense, on my view, the appropriateness of the application of a predicate is always an affair de jure. But sharing this view is not necessary to appreciate the distinction. If

you don't like it, for example, you can think that one kind of predicate is true in some non-normative, de facto (i.e. correspondence) sense of truth, and another is appropriate in a normative, de jure sense of appropriateness, and the argument still works, although at a more abstract level. In deference to this view, I will sometimes call a predicate itself de facto or de jure, although this can also be read as short-hand for "appropriate de facto" or "appropriate de jure".

Consider the proposition "Bruce is a cat." Now, suppose that Bruce is in fact a cat. When asked why I ascribe the predicate of cathood to Bruce, then, it seems sensible to say that predicating cathood of Bruce is appropriate in virtue of the kind of thing that Bruce is. He's a cat already, before anybody predicates anything of him, and therefore my application of the cat predicate is appropriate de facto (alternatively, Bruce satisfies the cathood predicate de facto). Bracketing metaphysical worries about this story, there is something that seems very sound about the idea that it's appropriate for us to call Bruce a cat because he's a cat. Cathood, then is appropriately predicated de facto. On the other hand, suppose I assert a proposition like this: "Jones is a citizen." One notices immediately that it's fruitless to quarry for some natural or metaphysical feature of Jones that makes it appropriate to ascribe the citizen predicate to her. Rather, citizenship is fundamentally a matter of what rules or norms we take to apply to Jones; it is a predicate that is appropriate in its application de jure rather than de facto.[3] Another way to put this is that

[3] I don't want to be too quick here, so let me say something more about de jure predicates. First, there will of course be some de facto judgments that justify to our treating Jones in a certain way, but the list of judgments that justify the treatment (and therefore make a judgment like "Jones is a citizen" true) cannot contain only de facto judgments. To see this, assume that [a...x] is a complete and correct list of the de facto judgments that appear in the explanation of our treating Jones in way w. There will always need to be another, further judgment (y) that because of [a...x] we *ought* to treat Jones in w. But this sets off an obvious justificatory regress. Call the form of a set of judgments like [a...y] (where [a...x] are de facto judgments, and y is a normative claim) a justificatory strategy. Now, given an instance of this justificatory strategy, we can always ask how y is justified. Usually, this question will be answered with the same justificatory strategy, that is, by appeal to some further set of de facto judgments [a'...x'], which will again require a normative claim y'. This will again be ultimately justified in the same way, and so on.

My suspicion – although I will not pursue it here – is that this is will also, ultimately, be the case with de facto judgments. It seems clear that the question whether to judge that Bruce is a cat cannot be *exhaustively* settled by the fact that Bruce is a cat. This means that, even if metanormative realism is true, the distinction between de facto and de jure judgments is one of degree and not of kind. There will always be some foundational normative or

citizenship is a normative *status*. With this distinction on hand, we can put things in the following way. Philosophers often think that autonomy is an appropriate predicate for beings like us de facto. I do not think that this is right. Rather, I think predicating autonomy of some being is always a de jure matter, or that being autonomous is having a normative status.

On this alternate view, what does it mean to say that Jones acts autonomously? Very abstractly, it means that we take Jones's acting as an instance of acting under certain norms, presumably because we view her or her action as meeting certain de facto or further de jure conditions under which this taking is appropriate. What, then, on this account, does it mean to say that Jones is autonomous? It means that we treat Jones's activities as expressing her relation to a practice of giving and asking for reasons. Of course, this does not involve denying that certain relevant de facto predicates apply to Jones, but what we are up to in calling her autonomous, on this view, is much more like what we are up to in calling her a citizen than it is like our calling her a *homo sapiens*. Autonomy is a normative status.

Specifying this conception in any detail would of course take much more space than I have available here. In what follows, I will fill in some details of the – broadly Hegelian – version of this view I favor, but I will do this mainly in the context of characterizing the explanatory payoff for a working theory in this general neighborhood.

In the first chapter, I'll describe some important work that the de jure conception of autonomy is poised to do in metaethics. There, I'll suggest that adopting the de jure conception promises a resolution to several major objections to the metaethical constitutivist project, which itself promises a highly desirable theory of moral reasons (constitutivist moral reasons are motivationally internal to agents, universal, and normatively non-arbitrary). In the second chapter, I'll

---

evaluative proposition (say, y") that we ought to adopt, whether in the same way that we adopt "Bruce is a cat", or in some other way. The challenge will always be to develop a justificatory strategy for such a grounding norm that doesn't appeal to some additional normative judgment, and therefore regress.

motivate a broadly Kantian account of autonomous action and provide some reasons to think that Kant's development of this theory runs into some damaging problems. The way to address these problems, I'll argue, is to modify Kant's account of autonomy de jure, in a way that leaves no room for the de facto conception of autonomy to do any work. Finally, I'll conclude the paper by returning to the metanormative question and consider a possible strategy for a metanormative justification of autonomy that does not rely on any supposed natural facts about freedom.

2 AUTONOMY AND KANTIAN CONSTITUTIVISM

There has recently been a good deal of interest in metaethics about the idea that some particular motivation is required in order to act. This is because if acting requires a certain motivation, and if acting is inescapable for beings like us, then the requisite motivation is also inescapable, and its correlate norm or norms could provide a hard-stop for the justificatory regress of normative reasons. Call the advocate of such a motivation a constitutivist.

In this chapter, I am interested in a fairly orthodox Kantian variety of constitutivism.[4] I will make my reasons for this clear in §2.1.1 below; for now, it is enough to appreciate the distinction between Kantian constitutivists and their naturalistic counterparts. Two commitments distinguish Kantian constitutivists as I of conceive them, from other proponents of constitutivism. First, Kantian constitutivists, following Kant, conceive of autonomy as both the constitutive *principle* of action and as a natural capacity for self-directed activity.[5] By contrast, other consti-

[4] I mean to exclude, for example, Velleman's current "kinda Kantian" view (2009). I take Korsgaard as the exemplar of Kantian constitutivism, although as often as I can, I will characterize her strategy in terms sufficiently general for it to reasonably count as a Kantian constitutivist strategy *sans phrase* (i.e. one that sticks pretty closely to Kant's own view, supposing that he was a constitutivist – see ns. 38 and 42 below).

[5] The naturalized conception of autonomy is strictly speaking a departure from Kant. For Kant, autonomy is a *metaphysical* capacity because he views it as a kind of causality that we cannot find in nature (cf. *G* 4:446). Contemporary Kantians tend to abandon the transcendental view in favor of the natural one. Thanks to Andrew Altman for pushing me on this point.

tutivists take it that the constitutive norm of action follows not from a principle, but from an intrinsic desire, disposition, or interest – although very often, these naturalistic motivations are, like the Kantian's principle, related to the capacity for autonomy.[6] Second, Kantian constitutivists conceive of autonomy in this way because they believe that standing under the jurisdiction of the autonomy principle is conceptually identical to having the autonomy capacity. From here on, I will refer to Kantian constitutivism as constitutivism *simpliciter*, although sometimes I will use the Kantian adjective to emphasize a distinction between Kantian and non-Kantian constitutivists.

In what follows, I discuss the import of two objections to constitutivism, which I will call the "shmagent" objection and the "no-metaethics" objection.[7] I take it that these objections are good ones, and that they point to a deep and serious problem for constitutivists. However, I will argue that they do not point to a problem with constitutivism as such; rather, they merely make it explicit that the conception of autonomy currently adopted by the constitutivists and their interlocutors cannot fill the role that the constitutivists assign to it.

This suggests the following disjunction (CAD): either there is something wrong with constitutivism because it requires autonomy to do work that it cannot do, or there is something wrong with the assumed conception of autonomy because it cannot do the constitutivist work. According to CAD, the shmagent and no-metaethics objections can be taken either as objections to the constitutivist project as such or as objections to the assumed conception of autonomy.

---

[6] See n. 1 above.

[7] For the shmagent objection, see Enoch 2006; for replies, see Velleman 2009, pp. 135-145 and Ferrero 2009; in response see Enoch forthcoming. An early version of this objection belongs to Railton (1997, pp. 73-79). For the no-metaethics objection, see Hussain and Shah 2006. Hussain and Shah officially target Korsgaard, although they suspect that their claim applies more broadly (2006, p. 265). I am thus taking them at their suspicion, not their doctrine, in targeting their argument at a slightly broader group of positions.

Call the assumed conception of autonomy *the standard conception*. On the standard conception, autonomy is a metaphysical or natural fact about agents. Many philosophers, the relevant constitutivists included, subscribe to the standard conception. I have already introduced an alternative conception of autonomy, by appeal to a distinction between predicates appropriately applied de facto and those appropriately applied de jure. On the standard conception, autonomy is appropriately predicated de facto. On the alternative conception, autonomy is appropriately predicated de jure. Kantian constitutivists are relatively unique in also subscribing to the alternative (de jure) conception, which they think follows directly from the standard conception.

My aim is to substantiate CAD by showing that the standard conception of autonomy, and therefore the Kantian justificatory strategy for the de jure conception that relies on it, is incompatible with constitutivism. I argue that the shmagent and no-metaethics objections follow directly from the attempt to deploy the standard conception in constitutivist argumentation, and that therefore the standard conception is not the right conception for constitutivism. I conclude that as long as there is no reason to think that the de jure conception of autonomy – given a justificatory strategy that doesn't rely on the standard conception – obviously runs afoul of the same objections (and I cannot see one), both disjuncts of CAD are live options, and so a viable de jure conception of autonomy is of significant metaethical interest.

2.1. Constitutivists and their Objectors

On the popular instrumental conception of practical rationality, a consideration counts as a reason for an agent to do something just in case the agent represents the consideration as the

means to getting something that he wants.[8] Following Bernard Williams, we can call reasons understood on the instrumental model internal reasons. The contrast class is external reasons, reasons that count as reasons for someone regardless of his desires.[9] The instrumentalist conclusion is that all practical reasons are internal. Call this view internalism and call its denial externalism.[10]

Many have thought that if internalism is true, moral philosophers must make a serious, even enterprise-undermining, concession to the skeptic. Namely, if no considerations (importantly, no moral considerations) count as reasons for someone independently of his representing them as the means to something he wants, then it is possible to imagine an ideally coherent immoral agent, who, because not disposed to take up any moral ends, is immune to rational criticism.[11] But internalism does not necessarily entail this consequence. One way to see this is to redescribe the skeptical worry as follows.

(1) If internalism is true, then there are no necessary ends.

(2) If there are no necessary ends, then there are no categorical imperatives.

(3) If there are no categorical imperatives, then all moral reasons are hypothetical.

(4) The antecedents of hypothetical imperatives apply contingently.

(5) Therefore, if internalism is true, all moral reasons apply contingently.

[8] This view is often (although perhaps wrongly) attributed to Hume. Smith 1987 is an important statement of instrumentalism that makes this attribution. For worries, see Millgram 1995. Hubin 1999 provides a brief, helpful literature review (p. 42 n. 1).
[9] Williams 1980
[10] For this paragraph, see Velleman 1996, pp. 170-72.
[11] See Street 2009, *pace* Gibbard 1990, p. 171, and 1999, p. 145.

Now, it is possible to grant internalism and to deny (1) by claiming that all rational agents, just in virtue of being rational agents, have some shared motivation or end – in other words, by adopting constitutivism.[12] If constitutivism is true, a certain class of reasons (reasons that appeal to the universally shared motivation) have all the attractive features of classical external reasons, but none of the unappealing metaphysical baggage traditionally associated with them.

According to constitutivists, just in virtue of acting intentionally, each agent is committed to taking some principle as normative because that principle is the constitutive norm of action. The idea of a constitutive norm of action can be understood by analogy: just as normative standards for building a house come out of understanding what is constitutive of a house, normative standards for acting come out of understanding what is constitutive of an action.[13] Insofar as one builds a house, one is beholden to the norms internal to the concept of a house, and thus to take certain principles – e.g. "put up a roof" – as normative. Likewise, insofar as one acts, one is beholden to the norms internal to the concept of action. Since we are, as Sartre put it, condemned to act, we are also "condemned" to be bound by certain norms.

Having appreciated the problems that arise for internalism regarding non-optional moral normativity, one can pretty easily see the attraction of constitutivism. If the constitutivist strategy is a good one, then the constitutive principle of action is a universal, non-optional, and non-arbitrary norm, and the moral skeptic can be shown that he is committed to this norm just in virtue of acting, which he cannot help but do. Constitutivists can thus concede that practical reasons apply only to those who have a certain motivation (i.e. can concede that internalism is true), but

[12] See Velleman 1996, p. 179ff. There are of course other options for resisting these consequences. For example, Korsgaard (1997) argues that the internalist argument either proves the internalist conclusion or else it describes who counts as a rational agent.
[13] Korsgaard 1999, p. 112

argue that since the relevant motivation is the motivation that makes one an agent in the first place, its application is not limited to agents of a particular temperament, and therefore the reasons that this motivation makes applicable are (say) quasi-external.[14]

Constitutivism, however, has its own share of apparent problems. Here, I am interested in what I take to be two particularly serious objections to constitutivism, which I will ultimately argue recommend the de jure conception of autonomy to constitutivists. What I am calling the shmagent objection to constitutivism has been most forcefully advanced by David Enoch.[15] The objection, put succinctly, is that it does not follow from the fact that we are condemned to act that we are condemned to care about acting, or, as Enoch puts it, "agents need not care about their qualifications as agents".[16]

Consider the moral skeptic, who does not take the fact that some action is immoral to give him any reason not to so act. Suppose that through constitutivist reasoning we convince him that some principle is a constitutive norm of action, that he cannot count as an agent and his movements cannot count as actions unless he adopts this principle. Enoch argues that there is no reason to think that this should this move him to act any more than did the putative fact that his proposed action was immoral. Why shouldn't we think, Enoch asks, that the constitutivist reasoning just shows the skeptic that he doesn't care about agency or action *either*?[17]

What I am calling the no-metaethics objection is advanced by Nadeem Hussain and Nishi Shah.[18] Hussain and Shah argue that constitutivism is consistent with a variety of metaethical positions, and this is a problem for constitutivism insofar as it is conceived as an alternative to, for example, non-reductive realism (and therefore externalism). According to Hussain and Shah,

---

[14] Velleman 1996, pp. 180, 199
[15] Enoch 2006 and forthcoming
[16] Enoch forthcoming, p. 2
[17] Enoch 2006, p. 179
[18] Hussain and Shah 2006

the relevant constitutivists have conflated the first-order normative question of which normative judgments we cannot help but make with the metaethical question of what it is to make a normative judgment in the first place.[19] The constitutivist claims that some inescapable normative judgment provides a back-stop for the regress of normative reasons, but even if this is true, constitutivists can't tell us anything about why such an inescapable judgment *itself* counts as normative, because constitutivism is simply not a theory of what normativity is.[20]

As the way I have just laid out the no-metaethics and shmagent objections indicates, part of my plan is to suggest that these complaints are related, and this can be seen here by noticing that the shmagent objection claims that the constitutivist does not have an answer to the metaethical skeptic, and the no-metaethics objection suggests that this is no surprise, since constitutivism is not really a metaethical theory in the first place.

2.1.1. Kantian vs. Naturalistic Constitutivism

We are now in a position to appreciate the reason I want to focus Kantian versions of constitutivism, rather than on constitutivism more broadly. The constitutivist is put on the hook by the shmagent objection to show that indifference to one's agency is not really an option. One standard constitutivist opening gambit is to characterize the shmagent challenge as analogous to this challenge: I might have a quasi-external reason to sacrifice a pawn if I am playing chess, but I can always give up playing chess. The problem, constitutivists will suggest, is that the game of

---

[19] *Ibid.*, pp. 269, 293

[20] Hussain and Shah also argue that Korsgaard's metaethical *constructivism* is inadequate to the task, but their claim comes apart from the claim that constitutivism leaves a certain normative fact to be explained, and that explanation is in principle possible on a variety of metaethical views. Even supposing that constructivism of the variety that Korsgaard (2003) or Street (2008) advocates could explain normativity, the point remains that constitutivism itself does not provide a competing explanation, and neither does it seem capable of arbitrating between candidate explanations. If Hussain and Shah are right, deciding between competing metaethical accounts, with or without constitutivism, will involve score-keeping on features unrelated to constitutivism. An early version of the worry about constructivism belongs to Darwall, Gibbard and Railton (1992, pp. 140ff).

agency – Intendo[21] – is crucially different than the game of chess, precisely because you cannot quit playing Intendo. This means that having the kind of motives that are constitutive of playing Intendo is also crucially different than having the kind of motives that are constitutive of chess, for the same reason. For games that you cannot avoid playing, there is no need for some further reason to play the game in order to be subject to reasons internal to the game. But the worry lingers: does this kind of unavoidability, to which constitutivists do seem entitled, actually *help*? Cannot the skeptic play Intendo "half-heartedly...under protest, without accepting the aims purportedly constitutive of it as [his own]"?[22]

The constitutivist strategy will be, roughly, that since efforts to challenge the capacities required for playing Intendo "cannot even get going without relying on them",[23] they are self-vindicating. The shmagent objection suggests that a putative agent needs a reason available outside of Intendo itself to be convinced that he should adopt the constitutive aims of Intendo (in the way, for example, being a chess player might give someone constitutive reasons to sacrifice a pawn, but these reasons only apply in virtue of further, non-chess reasons to care about playing chess in the first place). Since, however, there is no standpoint which the putative agent could occupy external to Intendo (here, as opposed to chess), any challenge of this kind will be dialectically unstable – the putative agent, just in wondering whether there is reason for him to care about his own agency, will belie his worry insofar as his just *asking the question* shows that he already does.[24]

[21] Elijah Millgram's *bon mot* (2009).
[22] Enoch 2006, p. 188
[23] Rosati 2003, p. 522
[24] Ferrero 2009, p. 310-11

This is because "the motives and capacities that constitute us as agents make critical reflection and action responsive to that reflection possible".[25] They are, in other words the motives and capacities that constitute our autonomy, or our ability to reflectively self-govern. Since the question "Should I care about agency?" requires in advance that the questioner does, in fact, care about evaluating various candidate norms, and since this is just what being an autonomous agent is all about, it seems that the skeptical question is either unintelligible or badly formed outside of the normative context of agency.[26] In this way, "our inclination to satisfy our autonomy-making motives and to exercise our autonomy-making capacities will be self-supporting".[27] This argumentative strategy deploys to the now-familiar thought that you can't quit playing Intendo, adding to it the further premise that you have to care about playing Intendo to count as playing it.[28]

There are three ways to understand the constitutivist's response here. Two of them make the response implausible; the third is not open to constitutivists who conceive autonomy in a purely naturalistic fashion. The first is to understand the response as characterizing a matter of fact about our contingent desires. In this case, even if it turns out that we all do care about the constitutive norms of action, this is not relevant to the question: that we *do* care is clearly insufficient evidence for the claim that we *should*.[29] The second is to understand the response as an appeal to dialectical necessity; this is the response that questions of the form "Should I care about being an agent?" are semantically defective. The idea is that neither agency, nor any alternative (e.g. shmagency) can be objectively correct, and so when you ask whether you should care about one of them, you are only asking half of a question. The very idea of "caring" invokes the crite-

---

[25] Rosati, op. cit.
[26] cf. Street 2008, p. 225; Velleman 2009, p. 145
[27] Rosati, op. cit.
[28] cf. Velleman 2009, pp. 142ff; Enoch forthcoming, pp. 8ff
[29] Enoch proposes this reading in Enoch forthcoming, pp. 13-17: "The but-you-do-care response is thus no response at all. It is utterly irrelevant."

rion, and that criterion will be implicit to agency (or shmagency, or whatever). Velleman says that "the idea that there must be a correct criterion to invoke, and that its correctness must be objective in a sense that invokes no criterion whatsoever...is nonsense".[30]

One line of response is Enoch's, who notes that the question "Should I care about being an agent?" does not *seem* defective in the same way that, for example, "Is the Empire State Building taller?" does,[31] and this seeming provides at least some *pro tanto* evidence that it's not.[32]

For my part, I don't know what to make of any of that; I do not know how one would decide whether the question is *in fact* semantically defective, and I am worried that any determinate answer to the question about semantic defectiveness is going to beg the question. The constitutivist can insist on principle that the question requires two arguments just in case it is *impossible* to ask the question from outside a framework with a constitutive norm. The skeptic can insist that it only has one just in case it is *possible* to ask the question in a framework-independent way. It is hard to see what reasons could be adduced that would provide a common framework for discussion here.

I don't think that further discussion can provide a way to get out of this stalemate, but I think we can carry it out a bit farther. For example, the skeptic might claim that, even supposing that the question *is* semantically defective, this does no damage to the his position. Consider the following argument form.

(6) Asking "Should I φ?" is semantically defective.

---

[30] Velleman 2009, 145
[31] This example is from Street 2008, 225
[32] Enoch forthcoming, p. 31

(7) Therefore, I cannot ask whether I should φ.[33]

(8) Therefore, I should φ.

It is not clear, the skeptic claims, why or how (8) is supposed to follow from (7). How can constitutivists who take this tack can earn a stronger conclusion than the conclusion that either I should φ without asking whether I should, or I shouldn't φ without asking whether I should?

One natural response from the constitutivist is that what the skeptic is assuming is that if her challenge fails, she must φ. And the constitutivist line of response above suggests that her challenge fails. But, the skeptic can respond, that is only true because the constitutivist assumes that "the skeptic" (the character in the story who asks "Should I φ?") need be understood as an *actual* character (e.g. the philosopher articulating the skeptical position). Alternatively, the philosopher articulating the skeptical position conceives of "the skeptic" as "the embodiment of a problem *we* face, because of *our* commitments."[34] So there is a question here about whether the skeptic must be understood as an actual agent (or shmagent), and thus as subject to a charge of dialectical instability. In other words, it is not clear that there being an agent (or shmagent) who can stably embody this challenge is to the point. And here again, either assuming or doubting that there is begs the question: the difference between the skeptic conceived as *an agent* (or shmagent) and the skeptic conceived as *a problem* is just the question of whether it is intelligible to conceive of this question in a framework-independent way.

Thus, even if you dislike the intuition that something's not *feeling* semantically defective is strong evidence for its not *being* semantically defective, you still have to explain why a claim

[33] Because, i.e., I wouldn't be asking a question.
[34] Enoch forthcoming, p. 20

about the inability to *ask some question* is supposed to entail a normative claim about doing something else (even when the something else is the object of pseudo-question). In other words, you have to adopt a norm for dramatizing the dialogue between the skeptic and the constitutivist. If the practical norm to which you appeal is not the supposed constitutive norm of agency, you're begging the question. If it is the supposed constitutive norm of agency, your victory is pyrrhic: in the dialectical setting in which we can rely on the claim that *x* is the constitutive aim of practical reasoning, we can, of course, show that *x* is the constitutive aim of practical reasoning.[35]

The third way for constitutivists to develop this response, and the best way, is to start by noticing that Enoch has not given us a determinate worry at all. What reason do we have to think that shmagents – nonagents who are very similar to agents but who lack the aim constitutive of agency (but not of shmagency)[36] – are *possible*?[37] Just *assuming* that they are is begging the question against the constitutivist. In other words, the constitutivist can respond that we have a reason to care about the constitutive aim of agency because as far as anyone has been able to show, *that's just what it is to have a reason*.[38]

Now, there is a significant challenge to this response. Namely, it threatens to close off the possibility of raising the question ("Should I φ?") from within alternative, competing frameworks *altogether*, determinate or not.[39] The question is about the status of the claim that what it is to have a reason is to care about the constitutive aim of agency. Is the claim supposed to be tracking a natural *fact* about reasons? If so, then we are back where we started; the question is supposed to be unable to be asked as a matter of natural fact, but, as Enoch's argument reminds us, natural

---

[35] cf. Enoch forthcoming, p. 27
[36] Enoch 2006, p. 179
[37] Velleman starts down this road in 2009, pp. 143ff, but does not get too far down it. Ariela Tubert also gestures at something like it in 2010, pp. 663-4. I will return to this line of response in §3.2.3.1 below.
[38] Enoch forthcoming, p. 37
[39] Enoch, op. cit.

facts are (at best) candidates for practical norms. This suggests, I think, that we need to understand the claim as tracking a fundamental normative commitment about our self-conception as reasoners. This is the function of the principle of autonomy for Kantians, and appeal to such a principle is, as far as I can see, the only way to enable the possibility of this line of response. We must opt for a form of constitutivism where autonomy is understood de jure, on pain of this constitutivist response running in a circle.

2.2. The Kantian Theory of Intentional Action and the Moral Law

I will return to the shmagent and no-metaethics objections in the next section, but before I do, I have to lay some action-theoretic groundwork. First, I am going to briefly illustrate the well-known claim that the belief-desire model of action is insufficient to capture the idea of intentional action. To see this, suppose that I want to make a lasagna. Suppose also that I believe that turning on the oven will help me satisfy this desire. On the belief-desire model, this is sufficient to account for my turning on the oven – an action, supposedly. Yet, I do not seem to have done anything that should qualify as *acting*. Rather, I just sat back passively and waited to see whether my beliefs and desires would get things going.[40]

The Kantian view of action is designed to capture the idea that acting is *active*, or that it involves an agent, by requiring that actions involve "principles" in addition to belief-desire pairs ("incentives").[41] When an agent acts on a principle, he has an evaluative belief with roughly the content that his desire and correlate non-evaluative beliefs are sufficient to get things underway, and he takes this belief as normative, or makes it count for himself as a rule for his action. So, in

[40] The standard statement of this (belief-desire) view is Davidson 1978, pp. 85ff. For discussion, see Velleman 1992, pp. 123-4 and Millgram 2009.
[41] "[Action is voluntary] insofar as it comes about according to maxims (maxims...[are] principles practically subjective..." (*LM*, 28:678).

order for my turning on the oven to count as an action on the Kantian model, I must first have an evaluative belief like this one: I have *reason* to turn on the oven.[42] This belief will count as a principle for me in virtue of my adopting it as normative, as a guide for action or a hypothetical imperative (to wit: If I want to make a lasagna, I ought to turn on the oven). In Kant's terms, when I do this, I make the principle my maxim. Having such a maxim is the mark that distinguishes action from other forms of behavior. In this way, action is itself a normative concept.[43]

A lot turns on this move, but first notice that making it is required for constitutivism to work. This is because constitutivism trades on various activities having internal norms, and compliance with these does not appear to be, strictly speaking, a matter of desire. For example, while it is perfectly coherent to want to make a lasagna but not to want to get up and turn on the oven, you cannot coherently take making a lasagna as your end and fail to be subject to the normative requirement that you turn on the oven (assuming that this is a norm internal to the activity of making a lasagna). It is adopting this constitutive norm as a maxim, along with taking up some end, that is supposed to motivate you to action on the constitutivist account – not some complex of desires.

Now, the introduction of principles to the action story yields two additional requirements, one for the Kantian about intentional action generally and another one for the constitutivist.[44] The action-theoretic requirement is *autonomy*: I must be the source of my principles, and this means that I must be the source of the principle by which I pick out principles (hereafter, the me-

[42] Buss 1999, p. 411
[43] See Korsgaard 2009, p. 81. I suspect that this idea is not unique to Kantian constitutivists. For example, it seems reasonable to read Velleman's discussion of an agent's involvement in his activity (2002) or Frankfurt's discussion of activity "attribution" (1977) as relying on a fundamentally normative distinction between action and mere activity.
[44] Sophisticated Humean views can accept the Kantian theory of intentional action but maintain traditional (skepticism-allowing) internalism. See Street 2008.

ta-principle).[45] If I am not, my will is heteronomous, which is another way of saying that I am not the determinant of my action. On the Kantian account, my will's heteronomy means that my bodily movement is no longer an action properly-so-called, and so if putative agents do not meet the autonomy criterion, the Kantian theory of action exhibits the problem it is designed to avoid. Second, the constitutivist requirement is that the meta-principle be *normatively non-arbitrary*. In order to get beyond the skeptical challenge attendant to traditional internalism, the constitutivist must provide a non-arbitrary answer to the question of why an agent should take some meta-principle to be normative. Without meeting the non-arbitrariness criterion, constitutivism exhibits the problem it is designed to avoid.

So the constitutivist project aims to provide an account on which the reason that we ought to take some particular rules for action as normative both has its source in the individual agent and is normatively non-arbitrary. Following Kant, we can call the entity that would satisfy these criteria a *categorical imperative*. And now the question is this. How is an agent supposed to get one?

Call the target categorical imperative "the moral law". Since adopting the moral law must be normatively non-arbitrary, there must be a reason (*R*) in virtue of which the moral law must be adopted. But apparently, *R* must be arrived at in a paradoxical way. On the one hand, *R* cannot be self-legislated, because the moral law is supposed to make a categorical normative claim, and because in order to self-legislate *R*, the agent must (in some sense) have the option not to impose it on herself. (If she had no such option, in what sense could she have *legislated* it, rather than merely endorsed it?) So if she self-legislates *R*, then *R* is hypothetical, and then so is the moral

[45] There is conflicting textual evidence about whether Kant himself takes autonomy to be required for an action to count as such. At least in the *Lectures on Metaphysics* it appears that he does: "...all practical propositions...must presuppose a freedom in me; consequently I must be the *first cause* of all [my] actions" (*LM*, 28:269; cf. McCarty 2009, p. 121). This exegetical issue to one side, the point is that contemporary Kantian constitutivists do – indeed *must* – believe this.

law.[46] On the other hand, the agent must be the ultimate source of the principle of her action, and so the source of *R*. This is because the source of *R* is determinative of her taking the moral law as normative, and as we have seen, if this source is not the (putative) agent, then in acting on principles determined by the moral law, she would not be acting properly-so-called at all. If *R* is not self-legislated, then adopting it is inconsistent with autonomy and therefore with the possibility of action. So *R* must be self-legislated. In sum, agents are required to impose *R* on themselves and also required not to.[47]

The constitutivist strategy is meant to resolve this paradox by inviting us to see the moral law – codified by the Formula of Universal Law – as a *description* of the very thing that enables an agent to act in the first place: autonomy, the capacity to give oneself a principle. The constitutivist syllogism thus begins with a natural or metaphysical claim (e.g. your will is free), then advances a conceptual identity claim (e.g. having a free will is identical to standing under the jurisdiction of the moral law), and finally concludes with a normative necessity claim (you necessarily ought to obey the moral law).

Kant adopts this strategy in the *Groundwork*, first arguing that we are free and then deriving the binding nature of moral obligation from our freedom.[48] The derivation is supposed to be valid because Kant thinks that standing under the moral law is conceptually identical to being

---

[46] Alternatively, *R* cannot be self-legislated, because if it is, it is not the reason for a self-legislated adopting of the moral law. Rather, it *is* the moral law. Thanks to Sebastian Rand for this formulation.

[47] Terry Pinkard calls this Kant's paradox (2002, p. 59). It is also, basically, the *Euthyphro* problem.

[48] See *G*, 4:447. Kant seems to reverse the order of this argument between the *Groundwork* and the second *Critique*. In the latter, Kant takes the fact that we experience obligations to imply that we are committed to viewing ourselves as free (since ought implies can), and therefore (by conceptual identity) as under the jurisdiction of the moral law. Whether the second *Critique* strategy, which Kant appears to have continued to endorse until the end of his life (cf. *MdS*, 6:383), is different than the *Groundwork* strategy in a metaethically interesting way is a complicated question, and I will not worry about it here, because in the sense relevant to my argument, the two strategies are structurally isomorphic. Both infer the bindingness of the moral law from a fact – in the *Groundwork* case, that we are free, and in the second *Critique* case, a phenomenological fact about the experience of obligation – by means of a conceptual identity, autonomy. I will focus on the *Groundwork* strategy, both because it is the one that modern constitutivists follow, and because it is unclear to me whether Kant's assertions in the second *Critique* are question begging in this problem context: "...we regard [the moral law] as *given*...not [as] an empirical fact, but the sole fact of the pure reason which thereby announces itself as originally legislative" (*KpV*, 5:31).

free from causal determinism (which Kant calls negative freedom).[49] In *The Sources of Normativity*, Korsgaard follows the *Groundwork* strategy, appealing to the fact that the will is free, and then asserting the identity of this fact with the content of the moral law.[50]

In her recent work, Korsgaard has shifted from Kant's strategy of identifying *freedom* with standing under the moral law to a strategy that involves identifying *acting* with standing under the moral law.[51] Although the content of her strategy has changed, the form – natural claim, conceptual identity claim, therefore normative necessity claim – has not. The argument appeals to a natural fact about rational agents, that we must exercise a capacity to act, and argues that since a description of this capacity is itself a candidate meta-principle (the moral law), agents find that they already give themselves the moral law, or are the source of their meta-principle as a matter of fact, and thus need not have another reason for adopting it. Just describing the problem, it is supposed, yields the solution.

2.3. The Role of Autonomy de facto in Kantian Constitutivism

Here, it becomes difficult not to equivocate, because the Kantian term "autonomy" now refers to three different things: the capacity, the law, and their supposed identity. Constitutivists are committed to both the standard (de facto) and the alternate (de jure) conceptions of autonomy, and further, to a judgment about their identity.

---

[49] "...[negative] freedom and the will's own lawgiving are both autonomy, and hence reciprocal concepts...for this very reason one...can at most be used only for the logical purpose of reducing apparently different representations of the same object to one single concept..." (*G*, 4:450).

[50] "[T]he will must have a law, but *because the will is free*, it must be its own law...Now consider the content of...the Formula of Universal Law. The [FUL] merely tells us to choose a law" (Korsgaard 1996, p. 98; my emphasis). See also Korsgaard 1989, p. 166; 1996, p. 94; 1999, pp. 111-112; 2004, p. 10; 2008, p. 12; 2009, pp. 107-109.

[51] Korsgaard 2009, pp. 75-6. This change is important for the constitutivist line because if the natural fact in question is freedom, then there is at least some conceptual space for the possibility that we sometimes do not act freely, and therefore, sometimes act in ways that are not subject to the authority of the moral law. There is no such conceptual room, Korsgaard thinks, if autonomy is conceptually identical to action itself – we always act *actly*. (Importantly, it is less clear why our *behavior* should always count as action.)

This can sound incoherent *prima facie*, because autonomy is either a natural property, or a normative property, but not both. However, the Kantian's claim is more subtle than this way of thinking suggests. The idea is that, in virtue of your being the kind of object that you in fact are, it is normatively necessary for you to judge that you stand under the moral law. For the Kantian, being autonomous de facto is identical to being *committed* to viewing yourself as autonomous de jure. In the remainder of this paper, I argue that this strategy is not a very promising one, both for substantive reasons (the identity claim is implausible) and for reasons of principle (constitutivist syllogisms that use natural claims as their major premise cannot provide the requisite normative necessity, because natural facts are insufficient to necessitate normative judgments). I also suggest that the shmagent and no-metaethics objections are *de re* objections to the fact that the capacity-law identity judgment goes unsupported.

To see this, let's return to my lasagna project, now stipulating that there is a constitutive, internal norm for the activity "making a lasagna" that can be put in a hypothetical imperative like the one I gave to myself: If you want to make a lasagna, you ought to turn on the oven. Then suppose that I fail to turn on the oven, and Megan comes in to investigate. "What's going on?" she asks. "I thought you were going to make a lasagna." I reply that I have changed my mind, and that I no longer want to make a lasagna. Now, although she might not like this, neither will she be able to accuse me of making any kind of obvious mistake: as long as making a lasagna is normatively optional, it is perfectly rational for me to adopt a position of skepticism about the normativity of making a lasagna.

Suppose, though, that she persists in her efforts to convince me by accusing me of making a non-obvious mistake. Namely, suppose she says "You can't fail to make a lasagna. Whatever you are doing, you are always already making a lasagna. Therefore, you can't be a skeptic

about the normativity of the constitutive aims of making a lasagna. Therefore, you should turn on the oven." In taking this line, Megan relies on a factual claim about whether or not being a lasagna-maker is an escapable feature of my natural condition. In this case, it is obvious that she is wrong. But suppose for the sake of argument that she is right. This still cannot get her what she wants as long as I can judge that there is another way to conceive my situation that has a similarly good or better explanatory relation to the facts.

What Megan wants is an argument that entails that I ought not judge myself to be, for example, making a shmasagna, which would involve all of the standard lasagna-making norms except the one currently under debate. Unfortunately, given the facts, judging myself to be a shmasagna-maker explains things better. (Indeed, it is only possible for her to criticize me if I can in fact fail to subject myself the relevant norm – fail, in this case, to feel any rational pressure to turn on the oven.) Since I don't feel the rational pressure that is constitutive of lasagna making, I can ask, in what sense am I supposed to be always already making a lasagna? Why should I think that I am incorrectly categorizing my activity?

Now, one might think that since the state of affairs settles the truth of Megan's inescapability claim, that claim suffices to settle whether or not I am committed to judge that I am a lasagna-maker. But it doesn't. Just as the fact that Bruce is a cat settles the truth of the claim "Bruce is a cat", the putative natural fact that I am a lasagna-maker settles the truth of the claim "Paul is a lasagna-maker", but the *further* claim that I am committed to judge that I am a lasagna-maker (the claim that I am a lasagna-maker de jure) can't be directly settled by any kind of natural fact, precisely because it's not a claim about how I *am*; it's a claim about what norms I am committed to judging myself to stand under, or about how I ought to be. This is a standard is-ought problem: natural facts are insufficient to necessitate normative judgments.

An analogous problem exists for Kantian constitutivism. The constitutivist suggests that we already have a reason to choose the moral law as our meta-principle because our capacity to act commits us to judge that we stand under the moral law. But, like Megan's lasagna-constitutivism, Kantian constitutivism supposes that full-stop normative necessity follows from the fact that we are supposed to be inescapably involved in some activity. And, barring further support, it does not, because it does not follow from the fact that we are condemned to exercise some capacity that we are condemned to *judge* that exercise under any particular description. Necessarily caring about one's autonomy is part of a conception of autonomy that is itself normatively optional.

What's more, even if the is-ought inference were unproblematic, there are substantive reasons to think that the minor premise of the Kantian constitutivist syllogism (that autonomy de facto is conceptually identical to autonomy de jure) is implausible. Consider the following, adapted from a famous case by Peter van Inwagen.[52] Suppose we found out that aliens have implanted tiny, heretofore undetectable controllers in all of our brains. The controllers determine which principles we will take as the guides for our actions. What we have found is that our conception of ourselves as autonomous de facto (as negatively free or as acting for reasons we give to ourselves) is systematically faulty. Would we thus *necessarily* stop treating each other as autonomous (i.e. as morally responsible)? As *acting* at all? Should we all *necessarily* become error theorists about moral responsibility and action? Surely affirmative answers to these questions are implausible. And if that is right, it is sufficient to highlight an intuitive gap between any de facto concept of autonomy and the de jure one. They don't *seem* conceptually identical.[53] This all suggests that the constitutivist story is implausible on two counts. First, the lasagna case indi-

[52] van Inwagen 1983, pp. 108ff
[53] A related point can be made by recalling to mind the order of explanation point above. See pp. 1-2 above.

cates that the presence of autonomy naturalistically conceived does not necessitate a normative commitment to autonomy. That case suggests that the *form* of the constitutivist argumentative strategy is implausible, because natural facts cannot directly entail norms. Second, the van Inwagen example shows that the absence of autonomy naturalistically conceived does not necessitate the absence of a normative commitment to autonomy, and the order of explanation point shows that the presence of a normative commitment to autonomy does not necessitate that there be *any* naturalistically describable autonomy capacity. These latter two points suggest that the constitutivist story is implausible in its *substance*; they are evidence that the autonomy capacity is not conceptually identical to being committed to viewing oneself as standing under the autonomy law. Together, these three arguments suggest that constitutivists need a different approach. Namely, they suggest that a good constitutivist argument will not claim that agents are as a matter of natural fact some way, but rather will show that agents are already *committed to judge* that this is the case.[54] The target will be facts about agents' normative *commitments*, not natural or metaphysical facts.

This, I think, is a version of the shmagent objection, which we can characterize, in the Kantian case, as the claim that the natural or metaphysical fact of autonomy cannot commit us to judge that autonomy de facto is conceptually identical to autonomy de jure, and so the kind of caring that autonomy de jure entails is not normatively necessary.

The no-metaethics objection arises next. To see this, let's return one last time to the lasagna case, now supposing that Megan has decided to press me further. The only way for her to proceed, as far as I can see, is to claim that the supposed conceptual identity has a special normative status. But then I can wonder *why*. For all that the constitutivist account says, this status can

[54] Understood in this way, Kantian constitutivism bears a close resemblance to more naturalistic varieties, which propose some fact about agents of the form "in fact, you *do* care about x". See Enoch forthcoming, pp. 7-17.

be explained by any metaethical account, including a non-reductive realist one. So the no-metaethics objection can be characterized as arguing that if we are committed to judging that autonomy de facto is conceptually identical to autonomy de jure, Kantians have not explained why or how.

One way to read both of these objections, then, is as objections to the supposed normative necessity of the conceptual identity of autonomy de facto and autonomy de jure. The claims, respectively, are that we are not committed to judge that these are identical (the shmagent objection) and that if we are so committed, why we are is left unexplained (the no-metaethics objection). These worries are only applicable to Kantian constitutivism if it relies on the standard conception of autonomy. It is *because* the Kantian attempts to deploy a conception of autonomy as a natural or metaphysical fact (the standard conception) that his justificatory strategy cannot help but presuppose an unsupported normative claim (that autonomy de facto ought to be identical to autonomy de jure). The upshot is that if autonomy is a constitutive norm of an inescapable human activity, this had better not be *because* that activity is inescapable.

2.4. Back to Action Theory

At the end of §2.1.1, I suggested that what constitutivists want to say to skeptics is that *what it is to have a reason* is to be committed to the constitutive aim of agency. I also suggested that this claim should not be supposed to be tracking a natural *fact* about reasons, on pain of falling back into the skeptic's trap. The Kantian approach to answering this question is to understand the claim about reasons to be tracking a fundamental normative commitment (about what it is to have a reason) of all rational agents, the moral law or the principle of autonomy. So the important metaethical question for Kantians is the question of how to justify this fundamental norma-

tive claim. They attempt to do this by appeal to autonomy understood as a metaphysical or natural fact about agents, and, in this chapter, I have suggested that this solution cannot work; it merely pushes the natural fact doing the work back a step, and so the skeptical challenge re-emerges.

What we need is a way to talk about the status of the autonomy norm whose justificatory strategy does not involve a premise that states a putative metaphysical fact, which will always have to be accompanied by some further normative claim in order to achieve justificatory closure. The way to start working toward an alternate strategy, I think, is to go back where we began – to the philosophy of action. That is what I will do in the next chapter.

3 ON BEING AN ALIEN CAUSE TO ONSELF

In the first chapter, I argued that Kantian constitutivists cannot provide a sufficient account of the metanormative foundation for autonomy. If I am right, then this has the unfortunate consequence that Kantian agents never act at all. What I want to assume for this chapter is that (free) action is possible, and that some explanation of freedom is available that can justify this assumption. So I will re-open the metanormative questions, and approach the question of autonomy from a different, action-theoretic angle. In particular, I will consider some popular accounts of what it means to act freely, and then adduce some problems with them that recommend a Kantian approach to autonomous action. From there, I will explain what I take to be the key problem with the Kantian action theoretic approach, diagnose its source, and finally advocate Hegel's action-theoretic solution to it. At the end of this chapter, I will return to the metanormative question about the justification of a criterion for having a reason, briefly indicating how a Hegelian approach to this question would work.

3.1. Three Stories of Free Action

In order to understand the Kantian account of free action, and why we should accept it, we will have to first glance at two other available stories. I will have to be quick, but the broad strokes should suffice for my purposes.

3.1.1. The Empiricist Story

On the traditional empiricist account of intentional action, actions are *caused* by mental states. This story, given its modern form by Davidson,[55] starts with the claim that doing something for a reason involves having a pro-attitude toward some end (i.e. a desire to do it), and believing that the proposed action is a means to that end. Together, these two constitute a reason for action, and such a reason, if it concludes in an action, is also the cause of the action. So let's call this the *causal* theory of intentional action.

The causal theory of intentional action strongly suggests a "causal sourcehood" view of freedom. Since actions are understood as causes, free actions, one is naturally inclined to think, are those that are somehow instantiated outside of the already-in-place causal chain. Unfortunately, this kind of counter-causal freedom also seems to be a fool's hope. Although from our subjective perspectives, it seems like alternative possibilities are open to us – to walk or drive, cook or go out, and so on – if an action is merely a case of causation, then the parts of us that cause action are themselves caused, and so on backward until we arrive at causes of our actions that are outside of ourselves.

So, given causalism about actions, freedom appears to be a kind of perspectival illusion. This is made vivid most easily by imagining that causal determinism is true,[56] but this is not the

[55] Davidson 1963, p. 3
[56] cf., e.g. van Inwagen 1983

only way to see it. Regardless of whether determinism is true, causal sourcehood seems incoherent: causal chains that are structured probabilistically or by random chance do not suggest the possibility of agent-guided actions any more than do deterministic ones. When we understand action as a fundamentally causal phenomenon, we also understand ourselves – our beliefs, desires, reasons, and ultimately activities – as explicable exhaustively in terms of the past and the laws of nature. We are confronted with the fact that, as Thomas Nagel nicely puts it, everything that agents do is part of a larger course of events that no one "does".[57] From a sufficiently objective perspective, we, *agents*, seem to be no different than anything else – no more free, anyway.

This suggests to many of us that something about the empiricist account of freedom has gone awry. Surely it is strange to think that we ought to – or *could* – understand ourselves and our actions as no different than rocks or quarks. Surely we should think that we have missed the explanatory boat here. The best way forward, I think, trends toward a more rationalist story about freedom.[58] We can understand it by appeal so-called Frankfurt-style examples.[59] In a typical Frankfurt-style example, we are asked to imagine an agent (Smith) whose thought processes are constantly monitored by an insidious neuroscientist (Black). Black wants Smith to behave a certain way – say, to buy an ice cream cone at time *t*. If Smith shows any indication that he will choose a different course of action, then Black will intervene to ensure that Smith buys the ice cream. But in fact, Smith just buys a cone in the normal way, leaving Black no need to intervene. Since Black does not intervene, it is plausible to think that Smith acted freely, despite the fact that because of Black's plan, Smith could not have done otherwise.

---

[57] Nagel 1989, p. 114

[58] There are others, which, for reasons I can't get into here, don't seem promising. Agent-causal paths seem to just give the problem a name, rather than attempting to solve it (cf. Chisholm 1964). There are also ideas about how causal indeterminacy "bubbles up" to human action (cf. Kane 2002), but, again, it is hard to see how random events in brains would count as a kind of control over our actions.

[59] Frankfurt 1971

This kind of intuition suggests that, perhaps, none of the empiricist commitments means that an agent cannot behave freely, since he can behave in accordance with reasons that he reflectively endorses, regardless of the fact that he didn't, metaphysically speaking, have any alternative. Unfortunately, if we maintain a strict empiricism, the intuition is also relatively easy to undermine by appeal to so-called manipulation cases. Imagine that when Smith was an infant, Black clandestinely implanted an electronic device into Smith's brain which ensured that Smith would choose to buy the ice cream at time *t*. Black's device ensures that Smith will bring about the action that Black intends, and only that action, at time *t*, and that Smith will reflectively endorse the action. In this example, Smith exhibits the sort of endorsement that is criterial for freedom. Smith's actions are undertaken in response to reasons he endorses. Nevertheless, Smith does not seem to be acting freely. This problem generalizes. Any chain of events and circumstances, *including* events and circumstances that satisfy more robust freedom criteria specified in terms of mental states, events, and their relations, seems to be – as a matter of principle – reproducible by a manipulative agent.[60]

3.1.2. The Rationalist Story

Getting out from under manipulation cases, I think, is best handled at an action-theoretic level. To see the way forward, recall some of the reasons I have already adduced for thinking that the empiricist account of intentional action is insufficient to the idea of *action*. The philosophical thought, again, is that actions are *active*, and given an exhaustively causal account of the relation of agents to their activities, we cannot capture the active quality of an agent's making something happen. Appeal to something like an agent's maxim or principle may help to complete the story of action. But if actions must be understood in this way, then we are owed a story about

[60] cf., e.g. Steward 2008, 144-5

the propositional character of actions. What I am calling the rationalist account of action focuses exclusively on this propositional aspect of action.[61]

According to the propositional model, actions are *expressions* of agent's relations to propositions. Imagine an army private, who, upon the approach of his sergeant, salutes his superior. On the empiricist model, the private's arm movement is understood as an action in virtue of its being causally issued by certain cognitive and conative mental states in the private, which are in turn understood as caused events. We can pick out the action as a particular event in a causal chain. On the propositional model, the private's arm movement expresses his deference to his sergeant. We can pick out the action by its expressing an agent's rational relation to a norm; the salute *expresses* the private's endorsement of the saluting-superiors norm.

On the causal account, the private's action must be explained in terms of matters of fact about the causal order of events. When we successfully explain what happened, we give an objectively true description of the relevant causal processes, conditions and so on. On the propositional account, the private's action is a matter of fact about his relation to a certain normative proposition. When we successfully explain what happened, what we do is give an objectively true description of the relevant norms; for example, the proposition "One ought to salute to express deference".[62]

The propositional account of intentional action suggests an alternate conception of freedom. On the rationalist view, when agents act, they track the world of normative facts. Freedom, from this perspective, is understood as a function of an agent's standing in the correct kind of relation to these normative facts, *viz.* being responsive to the reasons for action that the normative facts give her. The basic idea is that an agent who is unresponsive to these facts is not properly

[61] I call it "rationalist" for historical reasons (cf. Wollaston 1722), although it also has contemporary advocates who could reasonably be correctly classified as rationalist; cf. Kamm 1992. These references are due to Schapiro 2001.
[62] cf. Brandom 1979

situated to authorize her actions. In effect, her exercise of reason is so ill-conceived that it cannot confer legitimacy on her motives.[63]

Suppose doing *Y* is constitutive of doing *Z*, and that I mistakenly believe that doing *Y* is a way of doing ~*Z*. Then, when I authorize myself to be moved by the desire to *Y*, with the goal of ~*Z*ing, then there is a sense in which I have not authorized myself to do what I am now doing (~*Z*ing). Insofar as I have a general desire to do what it is reasonable to do, then when I am moved to act in ways that are, in fact, incompatible with satisfying that desire, there is a sense in which I have not really authorized my actions. We might say that my failure to correctly track the relevant norms has prevented me from exercising my power to guide myself by reasons, and so has prevented me from being free.[64]

This kind of account does not depend on causal sourcehood, and so manipulation examples might seem to lose their bite.[65] However, it is not clear that rationalists have really discharged the problem. The idea, on the reasons-responsive account, is that Smith counts as free as long as his buying an ice cream asserts a correct relationship to the normative facts about what he has reason to do, regardless of Black's causing him to do it. But it is not clear how this helps us to shake free of the intuition that Smith is still not *really* free, or, more generally, that *unfree* rational agents still authorize the motivating power of the desires that move them to act. There seems to be an important sense in which the power behind Smith's authorizations is not his own – it is Black's – and the reasons-responsive conception has not provided us with adequate therapy to dispel this intuition.

---

[63] cf. Buss 2008 for this and the following paragraph.

[64] For an example of a recent philosopher who adopts this view, see Wolf (1990). For example, "a person's status as [an autonomous] agent rests not only on her ability to make her behavior conform to her deepest values but also on her ability to form, assess, and revise those values on the basis of a recognition and appreciation of … the True and the Good."

[65] I will focus on two worries here, but it is important to realize that both epistemic (how do we know what the normative facts are?) and metaphysical (are the normative facts part of the "furniture of the universe"?) problems are traditionally associated with this account as well.

Kant's approach to this dilemma is to understand freedom as a necessary practical commitment of all rational agents, rather than as something to be proved or disproved theoretically. If our practical and theoretical commitments are so clearly compartmentalized, the thought goes, Smiths's circumstances cannot as a matter of principle undermine his freedom. By focusing on the practical standpoint, we find that the question "How will Smith act?" has no determinate answer for Smith until he decides how to act.[66] What is a simple fact from the perspective of a third-person observer is not a fact from the perspective of the agent herself. And so, as a practical matter, Smith must still determine for himself what to do.

3.1.3. The Kantian Story

According to Kant's theory of intentional action, both the empiricist and rationalist models action are correct, and both are incomplete: there are two aspects of intentional action, the causal aspect and the propositional aspect. When I – the agent – act, I stand in reflective endorsement relations to a normative proposition (I make it my maxim). So, on the one hand, acting is propositional: it expresses a proposition, evaluative belief, or principle which I endorse as a rule for my action. On the other hand, Kant also maintains that I instantiate actions causally, in the empiricist sense. While I am practically committed to understanding my actions as propositional, I am also theoretically committed to understanding them as causal.

Although on the theoretical side, because of the impossibility of causal sourcehood, I am committed to the view that my actions cannot be free; on the practical side, because of the Fact of Reason – the experience of subordinating our desires in the name of duty – I am committed to the view that my actions are autonomous. This is Kant's famous third antinomy,[67] which captures

[66] Hampshire 1983
[67] KrV A444-51/B472-79

the tension between the rationalist and the empiricist accounts of free action. Understanding the resolution of the third antimony is the key to understanding Kant's conception of freedom, and doing so requires that we understand Kant's conception of practical reason. This, again, I will explain by contrast to the empiricist and rationalist conceptions of practical reason.

Instrumentalism about practical reason is the view that all reasons for action are means-end reasons, and this view is naturally paired with the empiricist account of action. If instrumentalism is true, reasoning that concludes in an action constitutively involves having some ends (pro-attitudes), and some beliefs about the means to getting them, at which point causality takes over. And if the causal model is true, it must be the case that all that is required to reason practically are beliefs and desires. Instrumentalism accounts in an attractively simple way for the role of these attitudes in practical reasoning.

Traditionally, the rationalist account of action is correlated with a kind of "tracking" theory of practical reason. The idea is that, when we reason practically, we "track" facts of the normative order in the same way that, when we reason theoretically, we track the facts of the natural order. To borrow Tamar Schapiro's nice gloss on this difference, whereas on the empiricist model, action relates us practically to a theoretical world, on the rationalist model, action relates us theoretically to a practical world.[68]

On Kant's view, action relates us practically to a practical world.[69] Reasoning about what to do is not a matter of finding knowledge to apply in practice, as the rationalist model suggests; nor is it merely a matter of being caused to behave by one's conative and cognitive attitudes. For Kant, normative concepts are moves deployed to solve practical problems.[70] The most important, and general, practical problem in Kant's philosophy is presented by the Fact of Reason – by the

[68] Schapiro 2001, p. 97
[69] Schapiro 2001, p. 98
[70] Korsgaard 2003, pp. 322ff

fact that I am practically committed to my own autonomy. From the practical point of view, the facts relevant to my decisions cannot free me from the task of drawing my own conclusions about what I have reason to do.[71] This is what Kant calls *negative freedom*,[72] and the problem it presents me with is this: what should I do? The answer to that problem is Kant's conception of autonomy, or positive freedom: I should obey the moral law.

According to the Kantian conception of action, when I express my endorsement of the autonomy norm, or my commitment to autonomy, by acting in accordance with the moral law, I help to *construct* an essentially human (rational, noumenal) reality – to "confer on the sensible world the form of a whole of rational beings."[73] So freedom, for Kant, is constraint by norms that I give to myself. Free actions are just those that are undertaken to uphold my practical commitment to autonomy in the face of my theoretical commitment to determinism. Being free is just a kind of asserting one's commitment to freedom by willing actions that accord with the categorical imperative. Once again, the point is that what is a simple fact from the perspective of a third-person observer is not a fact from the perspective of the agent herself, and the assertion of the primacy of the first-personal standpoint *in actus* is how I can, for practical purposes, transcend the antimony between the practical and theoretical standpoints.[74]

Like the reasons-responsive view, on Kant's *self-constraint* view, the causal-objective point of view on agents and their actions does not threaten their freedom. In fact, Kant explicitly acknowledges this threat, perceives it as a threat, and offers positive, systematic reasons to get clear of it. He leans heavily on the perspectival distinction in order to assert his conception of autonomy – a practical, or what I have been calling *de jure*, conception of autonomy. I take it

---

[71] Buss 2008
[72] G 4:446
[73] KpV 5:43
[74] Velleman (2000) argues something like this -- that the freedom that counts where autonomous agency is concerned is epistemic freedom with respect to one's alternatives.

that Kant is urging here that we leave putatively *de facto* aspects of our metaphysical constitution out of our conception of autonomy (whether in the form of causal sourcehood requirements on freedom, or in the form of theoretical "tracking" relations in the epistemology of freedom).[75] Rightly, here, Kant insists that we cannot shake our commitment to thinking that we are free, any more than we can shake our commitment to the view that the physical universe is causally determined and that this excludes the possibility of freedom as a theoretical matter.[76]

But the worry is that Kant is simply asking us to *dissemble*. Just as we worry that so-called hard determinists are asking us to *ignore* one of our most fundamental practical commitments about ourselves – that we are the determinants of our actions – Kant is asking us to ignore the equally fundamental theoretical commitment that we are natural beings, mediate causes in a nexus of mediate causes. It is not clear that Kant's therapeutic program has managed to address the problem that manipulation worries represent away; rather, it might seem that he has simply thrown up his hands and claimed that we are forced to look elsewhere. The worry is that autonomy de jure, on Kant's account, is a consolation prize.

3.2. Autonomy and Contingency

I now want to suggest that what the failure of Kant's theory of freedom makes clear to us is that the problem of freedom constitutively involves a problem of self-understanding. On the one hand, we self-identify with our caused desires and drives in general. But we find this self-identification on unsure philosophical footing. From the point of view of theoretical reflection, it appears that what we identify with is not, after all, *us*. The desires and drives appear as what Kant called alien causes, determining our wills from the outside. On the other hand, we identify

[75] I argued in §2 that Kant does not, after all, manage to do this. But it does seem to be his suggestion.
[76] Nagel's *bon mot* is that "action is too ambitious" (1989, 114).

with our putative autonomy, which, because of what we've got in the first hand, we suppose to stand above our biological desires and drives, endorsing one or the other of them, although not any one in particular. There is thus an apparently irresolvable tension in our self-conception. We seem to think of ourselves as two incompatible things.

3.2.1. Reason and Desire

On the one side, empiricists identify practical reasons and desires – practical reasons are just causally efficacious desires (which become causally efficacious in some particular way in virtue of beliefs about how to satisfy them). On the other, rationalists identify practical reasons with theoretical reasons – practical reasons are just theoretical reasons for believing that some action is good, a conclusion arrived at in virtue of tracking the moral facts. Kantians do not advocate either identification.

The goal of the Kantian project is to successfully integrate both our commitment to autonomy (practical freedom), and our commitment to determinism (theoretical non-freedom). If practical and theoretical reasons were the same, and if they told us conflicting things about whether we are free, then the practical reasons would amount to theoretical reasons that stood in tracking relations to a worldly practical fact F, and the theoretical reasons would stand in tracking relations to a worldly empirical fact ~F. If we run practical reasons together with theoretical reasons, then, we are committed to viewing the world as one whose normative facts conflict with its empirical facts, and so a logically impossible claim (F & ~F) could be made truly of the world.

Neither can practical reasons be desires, on pain of the impossibility of autonomous action.[77] Since the determination of our will by desire is understood theoretically, we are committed to understanding desire-determined activity as unfree. So, Kantians think, free actions must be those in which we subvert desire, and we must do this by reasoning practically, where the constitutive aim of this activity is understood formally as asserting our moral autonomy (against our determined desires) by guiding our action in accordance with the moral law.

Let me say that another way. Since as biological beings, part of us – our desires and drives – are causally determined, acting on these desires cannot count as acting freely insofar as our biological bodies are subject to natural causality and thus are merely mediate causes in a nexus of mediate causes. From the experience of subordinating our desires in the name of duty, however, we find that we are practically committed to the idea that we have the capacity to act against these biological drives, and thus to act as a first cause rather than a mediate one. However, although we must thus be practically committed to freedom,[78] Kantian pure reason cannot account for it: freedom's theoretical determination is wholly negative.

This amounts to a kind of dualism about volition. In positing nature and freedom as distinct elements of an agent, Kant is committed to a view on which there are two distinct "springs" for action – one rational and one desirous.[79,80] Since duties and desires emerge respectively from these faculties, and since there is a sharp distinction in place between the faculties, what freedom requires is that one subvert one's desires and drives by reasoning practically.

---

[77] For Kant, all actions stem from desire, but free actions are ultimately (to borrow a distinction from Rawls) principle- rather than object-dependent. See Darwall 2006, p. 220.

[78] Compare KpV 5:3-4: The concept of freedom "constitutes the *keystone* of the whole structure of a system of pure reason, even of speculative reason... [whose concepts'] *possibility* is *proved* by this: that freedom is real, for this idea reveals itself through the moral law."

[79] cf. PdG §622

[80] I suspect that some Kantians will want to object here that the idea of "respect" is supposed to pave over this gap. But Kant is pretty clear that, "though respect is a feeling, it is not one *received* by means of influence; it is, instead, a feeling *self-wrought* by menas of a rational concept and therefore specifically different from all feelings of the first kind, which can be reduced to inclination or fear" (G 4:401n).

Kant is convinced by his views about the limitations of finite reason that we could never *know* that anyone has ever pulled this off successfully.[81] This means that on Kant's view free action is, in principle, undetectable by human agents; we wouldn't know it if we saw it. The problem is that, if we can't know what success looks like, we can't distinguish actions that satisfy – even *approximately* – the success conditions for freedom. And this seems to undermine any argument of his that might take our freedom as a premise.

What this means is that, because of our necessarily impoverished epistemic position, willing in accordance with the moral law requires me to conceive myself as a citizen of a *merely possible* world of pure rational legislators, the Kingdom of Ends.[82] This self-conception is (a) a further determination of the volitionally dualistic self-conception and (b) what gives rise to the charge of dissemblance put forth at the end of §3.1.3. It is, I think, what is fundamentally wrong with Kant's conception of freedom.

This is because, by Kant's own lights, if the Kingdom of Ends were actual, freedom would cease to exist. On Kant's account, although our desires and drives are what *prevent* freedom – the Kingdom of Ends – from being successfully actualized, they are also the condition of possibility for freedom in the first place.

Kant says that a "perfectly good will would...could not on this account be represented as *necessitated* to actions...[since] it can be determined only through the representation of the good." Hence, no imperatives apply to such a holy will; the *ought*, Kant says, "is out of place here".[83] In order for there to be an ought, the condition of possibility for the Fact of Reason and for our commitment to our own autonomy, there has to be a pathological will that is in constant

[81] G 4:407
[82] G 4:433
[83] G 4:414

separation from its own inner principle.[84] If the will's inner principle were to be realized, then there would be no autonomy: rational agents could not do what they ought to do, because oughts simply wouldn't apply to them. But conceiving of oneself as a member of the Kingdom of Ends is just conceiving of oneself as a holy will.

This is a controversial reading of the Kingdom of Ends, but I think it is the right one. To see this, consider the alternative reading. On this reading, the Kingdom of Ends is an ideal social situation in which every agent always regards his "private ends" as subordinate to the universal ends of which the KE is the systematic union.[85] So, although agents in the Kingdom always subordinate their private ends to universal ends, they nevertheless *have* private ends. Thus, there is no *prima facie* reason to think that oughts do not apply to these agents. Rather, we can think that oughts do apply to them, and (also) they always do what they ought to do. This conception is predicated on a distinction in the modal status of agents in the Kingdom of Ends and that of holy wills: holy wills do what they ought to do necessarily – and therefore, they are not subject to any oughts at all – and citizens of the Kingdom of Ends do what they ought to do as a matter of contingent fact – thereby allowing Kant to maintain that oughts still apply to them.

I am inclined to think of this modal distinction as a distinction without a difference. Consider the set of motivational structures of the citizens of a Kingdom of Holy Wills (H), and the set of motivational structures of the citizens of a Kingdom of Ends (E). (There is no reason to suppose that holy wills have *no* motivational structures; just that theirs are quite unlike ours.) It is clear that we cannot find a difference in H and E by appeal to the actions that they give rise to, nor by the results of their deliberation, however far those can be separated from the actions. Further, we can't find a difference in H and E by appeal to their principles of action. But then where

[84] I borrow this way of stating the problem from J.M. Bernstein.
[85] G 4:435

is the difference? We might find a difference in their *desires*, but not in the relevant desires – the principle-dependent ones that determine their actions and willings to be the actions and willings they are. It appears to me that any differences between the wills in H and the wills and E are beside the point.

If this is right, what it means is that on Kant's view, the constitutive aim of free action is to eradicate its own condition of possibility.[86] So freedom must have a (merely) regulative aim, a necessarily unactualized, *merely possible*, self-conception. Acting freely, on Kant's account, is acting like what you ought to be, but cannot.[87]

Freedom's status as a "mere ought" is also what gives rise to the dissemblance of Kantian moral agents. On the one hand, in acting freely, we are committed to the idea that this Kingdom of Ends cannot be actualized, because if it were, freedom would cease to exist. On the other hand, we are committed to thinking that our actions are actualizing the Kingdom of Ends – taking this as the constitutive principle of our actions amounts to our acting freely. So the dissemblance here can also be understood in this way: the truth about the actions of Kantian agents is that they are not aimed at what they claim to be. As a matter of fact, autonomous action is impossible – it cannot be actualized. I *know* this, and so I must dissemble; I must behave *as if* free action were possible. This suggests that there is something faulty or duplicitous in my self-conception. I do not understand myself *qua* agent correctly.

3.2.2. Freedom and Apriority

Kant's strategy for showing that we are committed to viewing our actions as free shows simultaneously that free action is impossible. According to Hegel's diagnosis, the problem starts

[86] cf. PdG §§618ff
[87] This objection is Hegel's. He is fond of saying that for Kant freedom is a "mere ought" (cf. EL §94Z).

when Kant conceives of an agent's volitional apparatus as comprised of two distinct faculties, subjective desire and objective practical reason. The complaint is that the failure of Kant's conception of freedom is a failure of self-conception. In order to escape a self-conception that requires this constant self-alienation, Hegel suggests a way in which the so-called subjective and objective aspects of volition can be *united*, rather than divided, in an agent. On Hegel's account, we can conceive of agents as free when their actions, *done for reasons given by their subjective desires*, also function to assert their freedom. All that it takes to actualize this criterion is that certain conditions are in place under which agents are able to understand the objects of their desires as the objectively correct ones.

Of course, Hegel agrees with Kant that the coincidence of subjective desire and objective practical law is not *necessary* for finite, embodied rational agents, in a naturally structured world. But, Hegel is quick to point out, this does not entail necessary non-coincidence, either. Rather, Hegel thinks, the coincidence must be achieved, and can be through a (cultural) project of self-understanding.[88] Hegel's way of stating this claim is that, "freedom is nothing but the knowing and the willing of substantial universal objects such as Right and Law, and the production of a reality that is adequate to them."[89]

In order to start unpacking that, and to see if it is of any help, we can note that for Kant, as well as for Hegel, following certain *rules* for willing – in Kant's case, the moral law; in Hegel's, the substantial rules of "Right and Law" – is constitutive of freedom. Both conceive of the world, understood as a context of autonomous action, as a kind of rule-governed practice. That

[88] Allen Wood (1990) calls this Hegel's ethics of self-actualization.
[89] IPH p. 63

is, both understand the world as a kind of practical problem for agents, structured according to and limited by certain rules.[90]

Following Tamar Schapiro's helpful suggestion,[91] we can flesh this idea out by appeal to Rawls's early essay, *Two Concepts of Rules*. Some rules, Rawls claims, are "summary rules" – rules that report on or summarize correct decisions made in particular cases, where these decisions are made, correctly, prior to the existence of the rule.[92] Rawls's example of a summary rule involves the assuming that act-utilitarianism is true, and imagining someone who is wondering whether he ought to tell his fatally ill friend that her illness is fatal, when that friend asks him. This can be decided, on utilitarian grounds, prior to the existence of a rule that tells him what he ought to do in this kind of circumstance. So a rule that, for example, one should not tell one's fatally ill friends that their illness is fatal can be arrived at by summarizing already-in-place correct decisions. The point to notice, Rawls insists, "is that someone's being fatally ill and asking what his illness is, and someone's telling him, are things that can be described as such whether or not there is this rule."[93]

Other rules, Rawls claims, are "practice rules" – rules that apply to an agent in virtue of his participation in a defined practice, whose rules are thus logically prior to any other potential evaluative constraints. Here, Rawls's example actions take place in the game of baseball. While it is possible for a man to run between two bags on the ground without the rules of baseball, he contends, it is impossible for him to "steal base".[94] So rules about whether one ought to steal base when a pitch goes wild can only be formulated from within the practice of baseball – they

---

[90] In Kant's case, the relevant rules are the a priori intuitions of the Transcendental Aesthetic and the categories, and in Hegel's, I think, they are the rules constitutive of the category of Actuality. I'm not going to worry much about either here.
[91] Schapiro 2001, pp. 100ff
[92] Rawls 1955, p. 19
[93] Rawls 1955, p. 22
[94] Rawls 1955, p. 25

can neither be assessed on merit nor can the actions they prescribe be *done at all* independently of or prior to the existence of the rules of baseball.

Schapiro argues that we ought to understand Kantian pure practical reason as a practice like baseball. The difference, of course, is that for Kant *all* actions are special cases of moves in the practical reason practice. Since the moral law (positive freedom) is the grounding norm of practical reason, it provides the rules of the game. Freedom is a general, formal rule by which all actions are governed, and this rule applies *a priori* and necessarily to all agents.

This is a very ambitious claim; it comes with the heavy argumentative burden that the advocate of such a completely general, *a priori* freedom practice demonstrates that whatever rules it specifies hold for all actions, i.e. demonstrates that constitutivism is true. When an agent assumes the role of a baseball player, to take a contrasting example, she always has the option to quit playing, and in fact the rules of baseball inevitably lead to the end of the game. Further, such an agent can dissimulate, acting under the auspices of being a baseball player, but *really* acting to satisfy some practice-independent end (e.g. in the case of point-shaving or a personal vendetta). The challenge for defending a *general* practice is that an agent needs to be shown unable to opt out of it. All of her actions must be "well-formed" in advance – this amounts to the challenge of showing that there are no available practice-independent self-conceptions (i.e. that one cannot do the equivalent of point-shaving).

As I understand Hegel's project, it is an attempt to save Kant's idea freedom as a kind of practice rule, his *de jure* conception of autonomy, although Hegel thinks that saving this will require a radical departure from Kant's theory of the will. In fact, Hegel thinks, problems of self-conception like Kant's are a necessary consequence of the supposed *generality* of the freedom practice. Hegel's claim is that any practice that is completely general is bound to be unsatisfying

in the way that Kant's putative freedom norm (the Kingdom of Ends) is, precisely because it must pit moral agents – who are always particular beings, acting from empirically determined desires in nature – against themselves. Kantian agents are divided; they are universal or objective in their reason, and particular or subjective in their desires. In Kant's theory of free agency, each side is intolerable from the perspective of the other. One's self-conception as a free agent is gained at the expense of one's self-conception as a biological being living in a certain culture at a certain point in history, and vice versa. Hegel's idea is that *any* conception of practical normativity as universal assures this dissatisfaction in advance, because it cannot, in virtue of its generality, be sufficiently attentive to the fact that if agents will anything at all, it must be something in particular, despite the fact that there's no particular thing that they must will.

The alternative to Kant's picture is to conceive of the rules of the freedom practice as themselves neither a priori nor general, but rather, as particular. Call a particular instantiation of a freedom practice a *culture*.[95] Now, unlike on Kant's view, a particular instantiated freedom practice (a culture) is not exhaustively derivable *a priori* from the constitution of rational agents – rather, it is determined by *embodied* rational agents developing rules in history – and so a given practice cannot be understood, as a matter of necessity, to function successfully as a criterion for free action. What then is the criterion of success for a practice? Hegel's claim is that it is the concept of freedom: the concept of freedom and a given practice function as each other's standards of correctness. Hegel's technical term for a successfully achieved freedom practice is the Idea [*Idee*] of freedom: the unity of the concept of freedom and its actuality in a practice. On this view, conditions for rational self-identification with the norms of a given culture are possible if these norms have the right kind of content and the right kind of epistemic status – if they can be

[95] Hegel calls it a state (*Staat*) for technical reasons that I won't get into, so to avoid the inevitable confusion this term brings up, I won't. See PR §33.

understood, from the perspective of individual agents who endorse them, as providing, on their own terms, a criterion by which actions can be judged successful. Hegel's claim is that none of the preceding accounts (empiricist, rationalist, or Kantian) can do this because each fails to conceive of willing individuals as both particular – embodied, and determined by causally efficacious mental states – and as universal – roughly, as the site of transcendental apperceptive activity. The Kantian account satisfies the latter criterion, and the two preceding accounts satisfy the former. I'll come back to this idea at the end of the paper; for now, I want to focus on Hegel's substantive claims about his idea of freedom.

For freedom to be actualized, on Hegel's account, the norms of a culture must be such that they do not prohibit individual agents from acting in accordance with their particular, subjective, causally determined desires. If the norms of a culture satisfy this criterion, then agents can understand some of their actions within the culture, regardless of the nature of their "springs", as the actions they *should* do, that is, as actions endorsed by their culture and satisfying their desires. They can understand their actions as upholding a freedom-enabling practice while at the same time acknowledging that they are determined. Regardless of the biological or socio-psychological origins of the relevant motivations, agents can correctly understand their actions as instantiations of freedom. They can have an integrated conception of themselves as a biological, temporally situated being that does not undermine their wholehearted endorsement of my activities – in this way, they can know that they are free without denying that part of what they am includes their natural situation.

In order to know I am free, Hegel thinks, I must not only be able to conceive of myself as free as a conceptual matter, but also I must be able to actually engage in self-determining action, and this means that I must be able to exert the determinations of my will in the world. Minimally,

Hegel thinks, this means that I must be able to acquire property.[96] If I can't rightly determine some part of the world outside my mind – my body, an ice cream cone – as *mine*, by exerting my will, then the supposed free determinations of my will (for example, my desires) cannot be made actual. As it turns out, there are a variety of further conditions that must obtain in order for the institution of property to be intelligible. So the idea that there must be property comes from the idea that there are wills, and this turns out to require that other institutions must be in place: property requires contract, which presupposes an idea of "wrong", and so on.

In the final analysis, certain concrete practices will be necessary in order for the world to be one in which there are (free) wills. These will be those that serve to make my freely willed actions intelligible as such. To take a quick example: if, in satisfying my desire for an ice cream cone, I am upholding a good market practice, my action is a good (free) one, in virtue both of the fact that it satisfies my subjective desire for an ice cream cone, and in doing so, also upholds a practice that is the enabling condition for the free action of all of the individuals over whom it has authority. Meanwhile, if I steal an ice cream cone, this can be a bad (unfree) action, despite the fact that I *wanted* to steal the ice cream, if it fails to uphold the rules of a practice that provides the condition of possibility for my understanding my actions as free.

Of course, understanding which desires we can rationally endorse is, on this account, also a function of the existing cultural norms. This means that we must be able to test the existing cultural norms against the idea of free action – action by a self-determining will. So, to continue our example, an exploitative market practice can be inconsistent with the possibility of free action for everyone that stands under it – and therefore, as a matter of contingent fact, for me. Such a practice occludes the possibility of objectively free action for some agents who stand under it, and thus cannot provide an appropriately public or objective standard by which actions can be

[96] PR §§41-44

understood as free. Agents under such a practice cannot know that they are free, and therefore, in the absence of an objective criterion of freedom, cannot undertake free action. We can understand this in terms of a pair of conditionals; given a successfully defined freedom practice, free actions are possible ($Fp \rightarrow \Diamond Fa$) and if any action is free, this implies a successfully defined freedom practice ($Fa \rightarrow Fp$). Of course, these conditionals are merely heuristic: freedom, on this account, is not a binary property, but rather one that comes in degrees as a culture progressively actualizes the concept of freedom.

This approach does not recommend blind advocacy of social or institutional rules, although it does invest a significant amount of credence in them. The onus is on the members of a given culture to act in accordance with its laws only until it becomes clear that the laws are stumbling blocks for free action. And as soon as the members of a culture can provide *reasons* for viewing the practice as unsuccessful, what they have shown is that the supposed practice rules are not, after all, the rules that actualize the concept of freedom.

That was a very terse explication of Hegel's idea of freedom. I want to move on quickly. But before I do, it is worth noting that, if Hegel is right, nothing like autonomy de facto could possibly be sufficient justification for autonomy de jure. In fact, thinking that it could is simply a species of logical error. Autonomy de facto is supposed to be a completely general and necessary natural feature of anything that we could appropriately count as an agent. But no general natural feature of agents could function as the sufficient justification for or the constitutive norm of a non-general freedom practice, simply because it could not provide a criterion by which we could tell free and non-free practices apart.

3.2.3. Problems

In §§3.1.1-2, I introduced a standard empiricist and a standard rationalist account of freedom. Both of these accounts, I suggested, run into serious trouble with determinism. On causal source accounts of freedom, determinism is obviously threatening: there seems to be no way to understand the universe as other than causally determined or else indeterminate in a way that does not conduce to human freedom. On rationalist, reasons-responsive accounts, agents authorizing the motivating power of the desires that move them to act can likewise seem to be threatened to be determined to so authorize. The worry in both cases is that the power behind an agent's actions or authorizations is not her own. The point of this kind of rumination is to suggest a kind of *bypassing* of the agent – in Frankfurt cases by an Orwellian neuroscientist, but more generally by an agent's causal makeup or social or biological history.[97] Kant's solution to this problem is to assert the fundamentality of the practical perspective in questions of freedom. Since from the practical point of view, none of the facts relevant to my decisions is intrinsically action-guiding, *I* must still decide what I have reason to do.[98] Kant's answer to this general practical problem – negative freedom – is to argue that (positive) freedom is a binding and completely general rule of individual practical reason – the moral law – and that by asserting my commitment to that law, I am free in the face of my theoretical commitment to determinism. In other words, for Kant, I am free when I do what I *should* do, rather than what I want to do.

Hegel characterizes agents on the Kantian model as *dissembling*: they act *as if* they were free, even though they know they are not. Kant does not so much address the bypassing threat, as try to give us reasons to think that we need not face it head on. But Kant's arguments to this effect do not work. Kant's argument that we merely *ought*, as a practical matter, to be free, is pre-

[97] The "bypassing" way of talking comes from Nahmias (forthcoming).
[98] Buss 2008

dicated on a conception of freedom whose achievement is as a matter of principle epistemically inaccessible to finite agents, leading to a self-conception on which we possess potentially a freedom that would cease to exist were it ever actualized. Hegel attempts to develop Kant's account of freedom as doing what I should do by situating the autonomy norms within a concrete culture. Agents in a culture can *know* these norms and can *know* by a process of internal critique if they count correctly as norms of freedom; they can also will in accordance with their subjective desires in ways that uphold the norms, and so can will in a way that they know is both subjectively and objectively correct. This is, for Hegel, free action: action that I know I would endorse regardless of whether I could do otherwise. It can seem counterintuitive that this kind of self-knowledge does so much of the work in Hegel's theory of freedom. Why should it matter that I know I am free? Can't I be free without knowing it? And so on.

It is helpful now to return to a manipulation case. Suppose Black is back, and up to his old tricks, manipulating Smith. Suppose that *y* is a set of actions $\{t_{y1},t_{y2},t_{y3}...\}$ where t_{y1} is the action that Smith performs at time t_1, and so on. Define $f(y_t)$ as a pointer function that runs over *y*, picking out the set member corresponding to the current time. Now, on the empiricist account, regardless of what $f(y_t)$ is, Smith cannot be free. The rationalist denies this claim, suggesting that if Smith rationally endorses $f(y_t)$, then he is free. But seems wrong, because it seems to require that the action at *t-1* was Smith's authorization of the desires that moved him to act at *t*, and so Smith is not responsible for $f(y_{t-1})$ unless he also authorized his authorization at *t-2*, and so on.

Kant agrees that this is a problem. In his version of this story, Smith knows about Black, and therefore he knows that he is not free. His solution is to focus on the fact that if Smith can procedurally demonstrate that his action is the *correct* one, practically speaking, then he doesn't need to worry about Black, because he can *know* that what Black is ensuring that he does, as a

practical matter, is what he would have done anyway. The point is that it does not matter whether Smith wanted chocolate or vanilla at *t*, as long as nothing that could ever have happened at *t-n* could have meant that Smith *should* have wanted, say, vanilla at *t*. The difference between Kant's point and the rationalist one is that Kant abandons the attempt to understand the freedom of a caused, phenomenal agent, and asserts that the freedom of a rational, noumenal agent *in spite of* the former. This is supposed to solve the rationalist regress problem by pointing out that, as a practical matter, an authorization at t-1 is *mine* from the practical point of view, regardless of the fact that it is not from the theoretical point of view.

This is unsatisfying. After all, all Kant has really done to solve the rationalist regress is to claim that as a matter of practical necessity I have to *take* an authorization as mine, even though I know that it is not. I can't understand myself as free, but I have to take myself to be regardless. Hegel's response here is really quite simple. Smith's job is simply to see Black not as a threat to his freedom, but rather as an enabling condition of it. If it turns out that Smith can understand $f(y_t)$ as both what he wants at a given time (which is true *ex hypothesi*), and as being what he should do, what he understands is that Smith *just is* Smith-caused-by-Black, and that this is *what Smith ought to be*. This amounts to Smith's understanding *himself* correctly – as Smith-caused-by-Black – and so as free.

This also suggests a problem with this kind of manipulation example; namely, that it seems to illicitly presuppose Black's freedom. In this case, the threat to Smith is that he is not free to determine his will because he must understand his will *as* Black's will. Because of the contrast between Black as we understand him and Smith as we understand him, Smith seems clearly not to be free. This can be misleading, because it is tempting to confuse a world in which there is a local threat to freedom (i.e., there is no problem in thinking that Smith isn't free and

that Black is, but then the lesson would be – don't be like Smith; get free), with one in which there is a supposed global threat to freedom. The right way to do the case would be to imagine a world in which Black also has a Frankfurt controller, Black*, who has Black** and so on back down the line indefinitely. When we conceive of this more developed picture, I think it is much easier to see how Smith might rightly come to consider his will as his own – after all, in this world, Smith's will is just what a will *is* – and therefore to take himself to be free.

3.2.3.1. The Capacity for Freedom, First and Second Order

One unusual upshot of Hegel's account is it appears that some or most humans are not as free as they could be, and for at least some of those, there is as a practical matter nothing that they can do about it. This does not seem too terribly wrong to me. In fact, it seems strange to imagine that, for example, very early members of the species *homo sapiens* were as free as modern individuals, for basically Kantian reasons. However, this fact does have the apparent consequence that we cannot correctly hold at least some of these individuals responsible for their actions.

A natural thing to want to say here is that all humans have the *capacity* for freedom, even if many of them have not (fully) actualized their potential for autonomy. Hegel, apparently, says something quite like this. He thinks that something is only free if it is free "in and for itself" – according to its concept (*an sich*) and self-consciously (*für sich*).[99] I have been discussing the second conjunct of this definition of freedom; it is why Kant's failure to provide concrete, epistemically available criteria for freedom is so troubling. However, with freedom *an sich*, two related problems arise for Hegel.

[99] EL §213

The first, and deeper, problem is that by introducing the idea of a capacity for freedom, we seem to sneak autonomy de facto in through the back door. The second problem is how we could be justified in holding any humans free only "in themselves" but not "for themselves" – that is humans with the capacity for freedom who are not free – responsible for their actions.

What is the difference between saying that all humans are free *an sich* and saying that they are autonomous de facto? What the Hegelian has to say here is that we are justified in understanding humans as having the capacity for freedom, not because of some putative metaphysical fact about them, but as a matter of fact about the norms of *our* culture. From a certain set of normative background assumptions, we are justified in thinking of and treating humans as possessing the capacity to be free, and the justifying reason is that doing so upholds the kind of practices that are the conditions of possibility for humans to *actually* be free.

As to the first problem, we need a distinction within the "capacity for freedom": the first-order capacity for free action, and the second-order capacity to get the capacity for free action. The first-order capacity is just *being free*, which, on Hegel's account amounts to correctly understanding yourself as free. The second-order capacity is the *capacity to get free*, or, the capacity to understand yourself correctly as free. Some humans, modern ones included, do not have the former capacity. Some humans, although no modern ones, do not have the latter. What is the difference? The difference is living in a culture where correct freedom norms (i.e. the norms of correct self-conception) are epistemically available.

It helps me to imagine a time-travelling philosophy professor from the present trying to convince a Teutonic peasant in the 2nd century BC that he, the Teuton, is free – that is, can do what he wants or choose not to. (This is also just fun.) I think it is reasonable to imagine that the professor's claim – "you are free" – would be basically unintelligible to the peasant, and further

that the reasons the professor gives to the peasant will also seem unintelligible, for Wittgensteinian reasons. Wittgenstein's famous point that "light dawns gradually over the whole"[100] suggests that the introduction of reasoning that is founded on radically different basic beliefs about *reasoners* will seem to us not to be reasoning at all. This because our reasoning and our picture of ourselves "proves itself everywhere, it is also a simple picture–in short, [we] work with it without doubting it".[101] In the peasant's case, the idea would be that the reasons he is free are not reasons *for him* until it becomes clear to him, for reasons internal to his own picture of himself, that that picture is untenable. The capacity to get free is just not available to him.

For those of us that live in the modern, liberal world, though, the capacity to get free is available to us – the reasons to believe that we are free, if there are any, are *our* reasons, there for the understanding. We have the capacity to get free, although some of us may be less free than others, and so not have identical capacities to act freely. So we are justified in thinking of and treating of modern humans as possessing the capacity to be free, at least insofar as they have the second-order capacity to have that capacity, and that just in virtue of their being what we (correctly) understand them to be. This is enough to ground attributions of responsibility of a kind that would be inappropriate for our peasant. (Think again about understanding Smith when we understood him on the empiricist picture.)

3.2.3.2. Back to Metanormativity, or: Are There Hegelian Shmagents?

Another worry about this account arises from the idea that we *correctly* understand humans to be free. If we are to understand a certain kind of community as succeeding or failing to provide the conditions of possibility for free action, we must have a criterion by which to judge.

---

[100] Wittgenstein 1969, §141
[101] Wittgenstein 1969, §147

According to Hegel, that criterion is the idea of freedom itself. However, the idea of freedom is only available in virtue of our culture's actually existing norms. It seems Hegel is arguing in a circle, justifying an idea of freedom by appeal to actual laws, and justifying those laws by appeal to freedom.

However, circularity is just what we should expect here.[102] Hegel's account claims to set the criterion for the correctness of practical reasoning; any such account cannot receive a noncircular justification without giving up its criterion-setting status. The question of whether a circular criterion-setting norm is correct cannot be settled by appeal to some criterion of correctness, on pain of begging the question. There is, however, the legitimate worry that many incompatible accounts can be self-justified in this way. How do we know that Hegel's is the right one? This is a particularly difficult kind of question to ask, because, as I have already suggested, you can't ask it without already assuming a criterion.[103]

I know of two approaches to moving forward. The first is to say that the question of a practical criterion is a matter to be settled by appeal to some theoretical criterion in the philosophy of action. Assuming this strategy, the metanormative foundation for the approach I have been advocating in this chapter is that Hegel's theory is consistent with having an integrated self-conception as agents, and that the satisfaction of this theoretical desideratum – which, I have also suggested, cannot be satisfied by non-Hegelian accounts of freedom – could reasonably be taken to motivate our acceptance of Hegel's practical claims.[104] Although I won't defend this claim here, I do not think that this approach will work, because I think that ultimately the norms of theoretical reason that one can deploy in the philosophy of action will end up needing to be

[102] Velleman 2009, pp. 141-2
[103] See §2.1.1 above
[104] This is roughly Velleman's tack in Velleman 2009; cf. p. 142.

grounded in practical reason.[105] Anyway, Kantians and their sympathizers (including Hegel) cannot adopt this approach for reasons of principle;[106] for now, that will have to be enough.

Alternately, we could call the very idea of a criterion into question. As I indicated above, this is in fact what Hegel does. I'll end the paper with a brief overview of his discussion of criteriality in the *Phenomenology of Spirit*, as a way of gesturing at his available, although almost never implemented, metanormative strategy.

The strategy involves showing that what we are supposing to be the *problem* of the criterion is actually a cipher for the *answer*.[107] Consider a schematic of the criterion problem, which, I think, is also the shmagent problem:

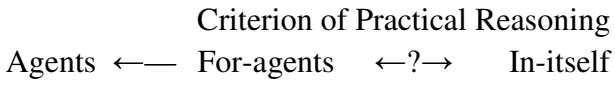

Figure 1. The Criterion Problem

What is the worry? It is that there might be a disconnect between the criterion of practical reasoning as it *appears* to us – who conceive ourselves as *agents* – and as it really is in itself. Even given an account that we have no substantive reason to doubt, we have a reason to doubt that we have got the right account in virtue of a general, in principle doubt that we are conceiving ourselves correctly. For example, we think we are agents, when *really* some of us are *shamgents*, and therefore we're all really some third thing. Therefore, this objection continues, the correct criterion of practical reasoning is whatever that criterion is in-itself – for agents and shmangents and whoever else – and it is in principle possible that this is different from the criterion as it appears to us, who might, as a matter of logical possibility, merely *think* we are agents.

---

[105] For Hegel's argument to this effect, see EL §§223-35.
[106] See, for starters KpV 5:119-121.
[107] PdG §§84ff

One of Hegel's great contributions to philosophy was to see past this problem. What he saw was that the idea of the criterion in-itself is *already* part of how the criterion appears for agents. The *distinction* between the criterion for-agents and the criterion in-itself is itself already part of the conscious activity of agents. If the criterion in-itself were in principle inaccessible to consciousness, then what we were representing when we took some rule to be the criterion of practical reasoning couldn't count as an attempt to represent the criterion of practical reasoning at all, which would be entirely beyond our representations.

In other words, in order to be *representing* something, our representing mental state has to refer to that thing, or be about it. Further, the mental state must be able to be evaluated against the thing with respect to properties like appropriateness and accuracy. But this evaluation would not be possible if the thing was not, also, already *in* our representing; we must be able to grasp the thing in order for it to be possible that we correctly or incorrectly represent it. This means, Hegel thinks, that the standard by which the thing (e.g., the criterion) can be judged must also be internal to our representing. The correct picture, according to Hegel, is this:

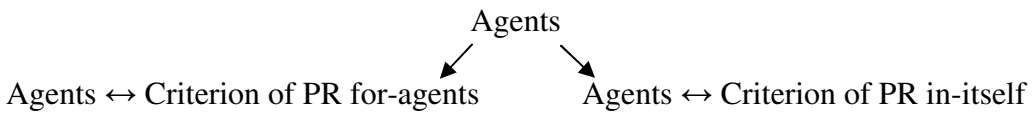

Figure 2. Hegel's Solution to the Criterion Problem

The question about the criterion in-itself, then, is not about the metaphysical status of the criterion – it's about the logical structure of my *activity* of representing it. And the question for a given a metanormative framework is whether, given the claims of the framework and their entailments, it's logically possible to get the criterion right. If I can't get it right, then I can't get it wrong either – the only way to appreciate the criterion problem as a genuine problem is to have a

self-conception that allows me to succeed or fail at providing an answer to it – and the idea is that only a being with a self-conception of a certain kind – i.e. a free self-conscious agent situated within a culture in which she can understand herself as such – could correctly understand herself as being in the position to do this, *or* as being in the position to fail at this in a way that could possibly be corrected.

Given this strategy, the metanormative foundation for the approach I have been advocating in this chapter is internal critique: the claim is that, on the Empiricist, Rationalist, and Kantian accounts, nothing that counts as practical reasoning, understood on the target account, could possibly be going on. I have already detailed Hegel's arguments about how Kant's view falls apart under this kind of internal critique in §3.2.1 above. The idea was that my self-conception as a free in Kant's sense, as a member of the merely possible Kingdom of Ends, means that as soon as I reason correctly (freely), I also fail (because I am a holy will). In other words, the criterion of practical reasoning for-agents could not be ever match up with the criterion in-itself, because if it did, then no one could ever successfully reason practically. This, to Hegel, suggests that Kant has got the wrong view about what it is to be an agent. In the *Phenomenology*, Hegel also attempts to show that what I called Empiricism and Rationalism suffer from this problem.[108]

Explaining Hegel's critiques of these views would take at least another paper, but what is important for my purposes is that this approach provides precisely the kind of account that I suggested was the best response to the shmagent worry in §2.1.1: claiming that we have a reason to care about the constitutive aim of agency because that's just what it is to have a reason. If Hegel is right, and practical reason conceived on the various non-Hegelian models cannot be actual-

[108] For example, empiricists identify practical reasons and desires, so the constitutive aim of practical reason is the satisfaction of desire. But this means that as soon as desire is satisfied, there's no practical reason (in the absence of the object of its constitutive aim, *what would it be*?). So on this model, nothing can count as the *actualization* of practical reasoning, because it is only practical reason as long as it has not satisfied its constitutive aim, and as soon as it does, it's not it anymore (cf. PdG §365). The attack on rationalism, I think, appears in PdG §§372ff.

ized, then this is *just* what he has shown – that the only coherent account of practical reason in town is the one that is underwritten by the Hegelian constitutive aim of action – self-understanding as the actualization of freedom. The strategy also addresses the worry about the status of the claim that what it is to have a reason is to care about the constitutive aim of agency. Having that reason is not a natural fact about reasons, it is the only account of having a reason that we've got that doesn't collapse under its own weight.

3.3. Being an Alien Cause to Oneself

Hegel agrees with Kant's idea of unfreedom as heteronomy – as having one's will determined by an alien cause;[109] his contribution is simply to note that, absent proper self-understanding, one can be an alien cause to oneself.

At the beginning of §3.2, I suggested that we re-conceive the problem of freedom as a problem of self-understanding. The job is to unify our descriptive conception of ourselves as biological bundles of caused desires and drives with our normative conception of ourselves as free. The apparent incompatibility of these two aspects of our self-conception provide us with a problem: to provide an account of a rule under which they can be unified in actuality. This rule, according to Hegel, is autonomy – being autonomous *de jure* is just the activity of willing our freedom, and knowing that by this activity, we are making both ourselves and the world free.

[109] G 4:446

4 REFERENCES

References: Kant and Hegel

References to Kant's works (except the first Critique) are given using an abbreviation for the title followed by the volume and page numbers of the *Akademie* edition of *Kants gesammelte Schriften* (Berlin, 1902–). References to the *Critique of Pure Reason* are given using an abbreviation for the title followed by the standard A and B pagination of the first (1781) and second (1787) editions respectively. The abbreviations and editions are as follows.

1. KrV: *Critique of Pure Reason*, Norman Kemp Smith, trans., (New York: St. Martin's, 1965)
2. KpV: *Critique of Practical Reason*, Mary Gregor, trans., ed., (Cambridge: Cambridge UP, 1997)
3. GMS: *Groundwork of the Metaphysics of Morals*, Mary Gregor, trans., ed., (Cambridge: Cambridge UP, 1998)
4. MdS: *The Metaphysics of Morals*, Mary Gregor, trans., ed., (Cambridge: Cambridge UP, 1996)
5. LM: *Lectures on Metaphysics*, Karl Ameriks and Steve Naragon, trans. eds., (Cambridge: Cambridge UP, 1997)

In all cases but the *Introduction to the Philosophy of History*, references to Hegel's works are given using an abbreviation for the title followed by paragraph numbers, with the *Anmerkungen* marked A and the *Zusätze* marked Z. References to the *Introduction to the Philosophy of History* are given using an abbreviation for the title followed by a page number. The abbreviations and editions are as follows.

1. PdG: *The Phenomenology of Spirit*, trans. A.V. Miller, (Oxford: Oxford UP, 1977)

2. IPH: *Introduction to the Philosophy of History: With Selections from the Philosophy of Right*, trans. Leo Rauch, (Indianapolis: Hackett, 1988)

3. EL: *The Encyclopedia Logic*, trans. T.F Geraets, W.A. Suchting, and H.S. Harris, (Indianapolis: Hackett Publishing Company, 1991)

4. PR: *Elements of the Philosophy of Right*, ed. Alan Wood; trans. H.B. Nisbet (Cambridge, UK: Cambridge UP, 1991)

I have occasionally slightly altered the above translations of Kant and Hegel, usually for reasons of consistency between and within the translations.

Other References

Brandom, Robert, 1979. "Freedom and Constraint by Norms", *American Philosophical Quarterly* 16:3, 187-96.

Brandt, Richard, 1979. *A Theory of the Good and the Right* (Oxford: Clarendon).

Buss, Sarah, 1999. "What Practical Reasoning Must Be if We Act for our Own Reasons", *Australian Journal of Philosophy*, 77, 399-421.

—, 2008. "Personal Autonomy", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/personal-autonomy/>

Chisholm, Roderick, 1964. "Human Freedom and the Self", in R. Kane (ed.) *Free Will* (Malden, MA: Blackwell, 2002), 47-58.

Darwall, Stephen, Allan Gibbard, and Peter Railton, 1992. "Toward Fin de Siècle Ethics: Some Trends", *The Philosophical Review* 101:1, 115-89.

Darwall, Stephen, 2006. *The Second-Person Standpoint* (Cambridge, MA: Harvard UP).

Davidson, Donald, 1963. "Actions, Reasons, Causes", reprinted in *Essays on Actions and Events* (Oxford: Clarendon), 3-19.

—, 1978. "Intending", reprinted in *Essays on Actions and Events* (Oxford: Press), 83-102.

Enoch, David, 2006. "Agency, Schmagency: Why Normativity Won't Come from What Is Constitutive of Action", *The Philosophical Review* 115:2, 169-198.

—, forthcoming. "Shmagency Revisited", in Michael Brady (ed.) *New Waves in Metaethics* (New York: Palgrave Macmillan).

Ferrero, Luca, 2009. "Constitutivism and the Inescapability of Agency ", in Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics*, vol. 4 (Oxford: Clarendon Press).

Frankfurt, Harry, 1971. "Freedom of the Will and the Concept of a Person", *The Journal of Philosophy* 68:1, 5-20.

—, 1977. "Identification and Externality", in Amelie Rorty (ed.) *The Identities of Persons* (Berkeley, University of California Press).

Gibbard, Allan, 1990. *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard UP).

—, 1999. "Morality As Consistency in Living: Korsgaard's Kantian Lectures", *Ethics: An International Journal of Social, Political, and Legal Philosophy* 110:1, 140-164.

Hampshire, Stuart, 1983. *Thought and Action*, (Notre Dame, Ind.: University of Notre Dame Press).

Hubin, Donald, 1999. "What's Special about Humeanism", *Noûs* 33:1, 30-45.

Hussain, Nadeem and Nishi Shah, 2006. "Misunderstanding Metaethics: Korsgaard's Rejection of Realism", in Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics*, vol. 1 (Oxford: Clarendon Press).

Kamm, Francis, 19992. "Non-consequentialism, the Person as an End-in-Itself, and the Signific ance of Status", *Philosophy and Public Affairs* 21:4, 354-89.

Kane, Robert, 2002. "Free Will: New Directions for an Ancient Problem", in R. Kane (ed.) *Free Will* (Malden, MA: Blackwell, 2002), 222-248.

Korsgaard, Christine, 1989. "Morality as Freedom", reprinted in *Creating the Kingdom of Ends* (Cambridge, UK: Cambridge UP), 159-187.

—, 1996. *The Sources of Normativity* (Cambridge, UK: Cambridge UP).

—, 1997. "The Normativity of Instrumental Reason", reprinted in *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (Oxford: Clarendon), 27-68.

—, 1999. "Self-Constitution in the Ethics of Plato and Kant", reprinted in *The Constitution of Agency* (Oxford: Clarendon), 100-128.

—, 2003. "Realism and Constructivism in Twentieth-Century Moral Philosophy", reprinted in *The Constitution of Agency* (Oxford: Clarendon), 302-326.

—, 2004. "Fellow Creatures: Kantian Ethics and Our Duties to Animals", The Tanner Lecture on Human Values.

—, 2009. *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford UP).

McCarty, Richard, 2009. *Kant's Theory of Action* (Oxford: Oxford UP).

Millgram, Elijah, 1995. "Was Hume an Humean?", *Hume Studies* 21:1, 75-94.

—, 2009. "Practical Reason and the Structure of Actions", *The Stanford Encyclopedia of Phi losophy (Summer 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2009/entries/practical-reason-action/>.

Nahmias, Eddy, forthcoming. *Rediscovering Free Will* (Oxford: Oxford UP).

Nagel, Thomas, 1989. *The View from Nowhere* (Oxford: Oxford UP).

Pinkard, Terry, 2002. *German Philosophy 1760-1860: The Legacy of Idealism* (Cambridge: Cambridge University Press).

Rawls, John, 1955. "Two Concepts of Rules", *The Philosophical Review* 64:1, 3-32.

—, 1980. "Kantian Constructivism in Moral Theory", *The Journal of Philosophy* 77:9, 515-572.

Rosati, Connie, 2003. "Agency and the Open Question Argument", *Ethics* 113:3, *Centenary Symposium on G.E. Moore's "Principia Ethica"*, 490-527.

Schapiro, Tamar, 2001. "Three Conceptions of Action in Moral Theory", *Noûs* 35:1, 93-117.

Smith, Michael, 1987. "The Humean Theory of Motivation", *Mind*, New Series, 96:381, 36-61.

Steward, Helen, 2008. "Moral Responsibility and the Irrelevance of Physics: Fischer's Semi-Compatibilism vs. Anti-Fundamentalism", *Journal of Ethics* 12:2, 129-145.

Street, Sharon, 2008. "Constructivism about Reasons", in Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics*, vol. 3 (Oxford: Clarendon Press).

—, 2009. "In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters", *Philosophical Issues* 19:1, 273-298.

Tubert, Ariela, 2010. "Constitutive Arguments", *Philosophy Compass* 5:8, 656-666.

van Inwagen, Peter, 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

Velleman, J. David, 1989. "Epistemic Freedom", *Pacific Philosophical Quarterly* 70, 73-97.

—, 1992. "What Happens When Someone Acts", reprinted in *The Possibility of Practical Reason* (Oxford: Clarendon), 123-143.

—, 1996. "The Possibility of Practical Reason", reprinted in *The Possibility of Practical Reason* (Oxford: Clarendon), 170-199.

—, 2002. "Introduction", in *The Possibility of Practical Reason* (Oxford: Clarendon), 1-31.

—, 2009. *How We Get Along* (Cambridge, UK: Cambridge UP).

Williams, Bernard, 1980. "Internal and External Reasons", reprinted in *Moral Luck* (Cambridge, UK: Cambridge UP), 101-113.

Wittgenstein, Ludwig, 1969. *On Certainty* (Malden, MA: Blackwell). References given by paragraph number.

Wolf, Susan, 1990. *Freedom within Reason* (Oxford: Oxford UP).

Wood, Allen, 1990. *Hegel's Ethical Thought* (Cambridge, UK: Cambridge UP).

Wollaston, William, 1722. *The Religion of Nature Delineated*, excerpted in D.D. Raphael (ed.), *British Moralists: 1650-1800,* vol. I, (Indianapolis: Hackett Publishing, 1991), 239-58.