

Kant's Conception of Autonomy in Two Objections to Metaethical Constitutivism

Paul Tulipana

To be presented at the APA Pacific Division Conference, April 2011

ABSTRACT: This paper suggests that two recent objections to metaethical constitutivist views in the Kantian vein can be traced back to the conception of autonomy that those constitutivist theories employ. On this conception, autonomy is conceived both as a metaphysical fact about agents, and, for that reason, as a constitutive norm that categorically applies to them. I suggest that David Enoch's *shmagent* objection can be understood as the complaint that autonomy as a metaphysical fact does not entail autonomy as a law, and that Shah and Hussain's *no-metaethics* objection can be understood as the complaint that if autonomy as a metaphysical fact entails autonomy as a law, Kantians have not told us why. This diagnosis suggests that it is open to those sympathetic to Kant's project to concede that these objections hold given a certain conception of agency, but need not therefore despair of the Kantian moral desiderata.

§

In this paper, I offer a brief, ground-up diagnosis of two contemporary objections to Kantian constitutivism,¹ which I will call the *shmagent objection* and the *no-metaethics objection*.² I will not directly defend the view that either of these objections obtains, although I think that they both do. Rather, I will merely assume that the objections are good, and then offer a structural explanation of why it makes sense that this is the case. Namely, I will suggest that if these objections obtain, it should come as no surprise given the conception of autonomy that Kantian constitutivists appear to employ.

This conclusion suggests a program of metaethical argumentation about the cash value of which it is worth saying something. For starters, if I am right, then even if the objections are good, they are hardly conclusive evidence against the hopes of the Kantian metaethical project generally. Rather, the following disjunction is true: either they are evidence that something is wrong with the idea of the constitutivist project or they are evidence that something is wrong

¹ The "Kantian" flag means to indicate a more-or-less orthodox Kantian variety of constitutivism, potentially excluding, for example, Velleman's current "kinda Kantian" view (2009). I will take Korsgaard as the exemplar of the kind of constitutivist I have in mind, although I think that many of the details of her view (outside of those that mark her as a more-or-less orthodox Kantian) are irrelevant to either the objections under discussion or to my way of diagnosing the source of their complaints. So, 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*. Hereafter, I will refer to this kind of view as "constitutivism" *simpliciter*.

² For the *shmagent* objection, see Enoch (2006); for replies, see Velleman (2009), 135-145 and Ferrero (2009); in response see Enoch (forthcoming). At least a proto-version of this objection belongs to Railton (1997), 73-79. For the *no-metaethics* objection, see Hussain and Shah (2006). Hussain and Shah officially target Korsgaard, although they state explicitly that they suspect that this claim applies more broadly (265). I am thus taking them at their suspicion – not their doctrine – in targeting their argument at a broader group of metaethical positions.

with the conception of autonomy operative in it. It is thus open to those sympathetic to Kant's project to accept that these objections obtain on certain brands of constitutivism (including, I think, Kant's own), but we needn't therefore despair of the Kantian moral desiderata.

Although it is my view that there is in fact something importantly wrong with the Kantian conception of autonomy – namely, that on this conception autonomy is understood as a metaphysical fact about agents – the argument here is several steps removed from directly claiming anything so robust as that. If I am right that the shmagent and no-metaethics objections to constitutivist views can be understood as following from the Kantian conception of autonomy, this is some evidence that the Kantian conception of autonomy is at odds with the role that the concept of autonomy plays in the constitutivist strategy. And insofar as there is reason to think that the conception is incapable of filling a philosophically attractive role for the concept to play, this is in turn evidence that something is wrong with the conception itself.

I.

The constitutivist line runs as follows: just in virtue of being an agent at all – that is, just in virtue of acting intentionally – each agent is supposed to be bound to take some principle as normative, because that principle is a constitutive norm of intentional action.³ The idea of a constitutive norm can roughly 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 (Korsgaard, 1999, 112). 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. Thus, since we are, as Sartre put it, condemned to act, we are also "condemned" to be bound by certain norms.

If this strategy works, the payoff is significant. Namely, constitutivist accounts promise to resolve two traditionally opposed and highly desirable criteria for a theory of moral reasons. On the constitutivist view, moral reasons are internal to each agent, and therefore satisfy the requirements for existence-internalism about reasons while also being non-optional, non-arbitrary, and so on.⁴ Constitutivism thus promises to provide a response to the moral skeptic without appeal to so-called external reasons, since on it, the advocate of universal moral normativity can point to moral reasons of a non-mysterious kind that already apply to the skeptic just in virtue of his acting intentionally. Of course, for this to work, constitutivists must account for the way in which the normativity of these reasons stems from the necessity of intentional action. For Kantians, autonomy is supposed to be the source of that normativity.

What I am calling the shmagent objection to constitutivism has been most forcefully stated by David Enoch (2006). The thought behind this objection is simple: it does not follow

³ Korsgaard thinks it is a principle (2008), 5 fn. 9. For other constitutivists (perhaps less orthodox Kantian ones) the constitutive thus-and-such is an aim, a motivation, or whatever. This makes a difference, but I think that Korsgaard's line is the most faithful to Kant's, so I will address these remarks to the idea of a constitutive principle.

⁴ For the relevant sense of existence-internalism and externalism about reasons, see Williams (1980).

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" (forthcoming). According to shmagent objection, if the constitutivist has got a convincing answer to the first order normative ("why-care-about-morality") question, she is nevertheless lacking a convincing answer to the second order ("why-care-about-being-an-agent") question, and so skepticism resurfaces at the metanormative level. What I am calling the no-metaethics objection is due to Nadeem Hussain and Nishi Shah (2006). The complaint here is that the constitutivist position is metaethically uninformative, or that it is consistent with at least a variety of metaethical positions, importantly including the non-reductive realism (and hence the reasons externalism) that constitutivism is partially designed to disallow. Part of my plan is to suggest that these complaints are related, and this can be seen at a high level of abstraction by just 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.

II.

Now I have got to lay some action-theoretic groundwork, the import of which should become clear shortly. First, I am going to briefly advocate the well-known claim that (what is often thought of as) the Humean theory of intentional 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 in order to make a lasagna, I have got to turn on the oven. In other words, I have a desire and a belief about how to satisfy that desire. On the Humean view, this is sufficient to account for my turning on the oven – an action, supposedly.⁵ The Kantian's complaint with this account is that on it, I (the agent) do not seem to have *done* anything. Rather, I just sat back passively and, in Elijah Milgram's nice phrase, waited to see whether my desires actuated my body (2009).

The Kantian view of action is designed to capture this idea, that acting requires an *agent*, by requiring that intentional actions involve principles in addition to belief-desire pairs – "incentives" in Kant's language (cf. Korsgaard, 2004, 6ff). Acting on a principle (and so acting properly so-called) requires that an agent has an evaluative belief with roughly the content that her desire and correlate non-evaluative beliefs are sufficient to get things underway, and that she takes this belief as normative, or makes it count for herself as a rule for her action (cf. Buss, 1999). So, in order for my turning on the oven count as an action on the Kantian model, I must have an evaluative belief like so: If you want to make a lasagna, you ought to turn on the oven. When I adopt this belief as a guide for my action, it becomes my principle, and having such a principle is the mark that distinguishes full-blooded action from other forms of behavior.

A lot turns on this structural move, but first notice that making it is required for the constitutivist theory 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

⁵ The traditional statement of this position is Davidson (1978), 86. For discussion, see Velleman (2006), 197ff.

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 you understand that this is a norm internal to the activity of making a lasagna. It is this constitutive norm that is supposed to motivate you to action, on the constitutivist account, and not some additional desire.

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.⁶ The action-theoretic requirement is *autonomy*: I must be the source of the principle by which I pick out principles (hereafter, the meta-principle). If I am not, Kant says, my will is heteronomous, which is another way of saying that I am not the determinant of my action. On the account given above, this means that my action 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 *normative non-arbitrariness*. In order to get beyond the pale of traditional reasons internalism and its attendant skeptical challenge,⁷ 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.

I have just described the desiderata of the constitutivist project. The job is to offer 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 (existence internalism about reasons) 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 Kantian's target categorical imperative "the moral law". Since adopting the moral law must be normatively non-arbitrary, it appears that there must be a normative reason in virtue of which it, rather than some other categorical imperative, must be adopted. But it also appears that this reason (ϕ) must be arrived at in a paradoxical way. On the one hand, ϕ cannot be self-legislated, because the moral law is supposed to make a categorical normative claim, and because in order to self-legislate ϕ , the agent must (in some sense) have the option not to impose it on herself. So if she self-legislates ϕ , then ϕ is hypothetical, and then so is the moral law. On the other hand, the agent must be the ultimate source of the principle of her action, and so the source of ϕ . This is because the source of ϕ 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 an agent properly-so-called at all. If ϕ is not self-legislated, then adopting it is inconsistent with autonomy and therefore with the possibility of intentional action. So ϕ must be self-legislated. In sum it seems that agents are both

⁶ Sophisticated Humean views can accept the Kantian theory of intentional action but maintain traditional (skepticism-allowing) motivation internalism. See Street (2008).

⁷ The challenge is that on a complex of views that includes existence internalism about reasons and theory of practical reasoning founded on a belief-desire model of intentional action, all reasons for acting are contingent, based on whatever occurrent motivations an agent happens to have. So, then, ideally coherent moral skeptics are intelligible and immune to rational criticism.

required to impose ϕ on themselves and also required not to. Following Terry Pinkard, we can call this Kant's paradox (cf. 2002, 59).⁸

The constitutivist strategy is meant to resolve this paradox by inviting us to note that the moral law – codified by the Formula of Universal Law – is just a *description* of the very thing that enables an agent to act intentionally in the first place: autonomy, or a kind of agential faculty in virtue of which the agent has the capacity to give herself a principle. This is Kant's Fact of Reason argument (1997, *Ak.* 531ff), echoed by Korsgaard like so:⁹

"[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." (1996, 98; my emphasis)

Agents are autonomous, have the capacity to give themselves a principle, and 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.

III.

Here, it becomes difficult not to equivocate, because the Kantian term "autonomy" now describes two different things. And indeed, the success of the constitutivist move turns on its doing so, on the ability to describe both the constitutive norm of intentional action (i.e. the moral law) and a capacity or faculty that agents have as a matter of metaphysical fact (i.e. the fact of reason) as autonomy. The constitutivist argument goes: because agents are (metaphysically) condemned to exercise the autonomy capacity,¹⁰ they are condemned to adopt its internal standards (the autonomy law) as normative. Because autonomy-fact, autonomy-law.¹¹

This adverts to a problem in the operative conception of autonomy. The problem is a standard is-ought one, namely, that it is unclear why autonomy as a metaphysical fact (autonomy-fact) should entail autonomy as normative fact (autonomy-law). In the remainder of this paper, I suggest that this supposed entailment is of dubious merit, at least barring a significant amount of further support. I also suggest that the shmagent and no-metaethics objections are *de re* objections to the fact that this putative entailment 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

⁸ Pinkard's characterization is notably vaguer than mine is, so any fault in my analysis lies solely with myself.

⁹ See also Korsgaard (1989), 166.

¹⁰ Strictly speaking, for Kant, because we are committed to believing this.

¹¹ At least this seems to be the suggestion. Compare, for example, Korsgaard (1996, 94); (1999, 111-112); (2004, 10); (2008, 12); (2009, 107-109).

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." In taking this line, she is relying on two sets of claims – one about the substantive internal standards of making a lasagna, and another about whether or not making a lasagna is escapable. Although in the case of the lasagna, it appears she is wrong on the latter count, suppose that I grant her both claims. In doing so, am I conceding that my behavior is instrumentally irrational? No, because it is open to me to say this: "Fine, I am always already making a lasagna, but I don't care. I don't take that as a reason to be motivated to turn on the oven." Here, I am suggesting that she is making a *category mistake* in appealing to some putative metaphysical fact, or to a description of my situation, and in supposing that it follows from this that some principle should be normative to me. The problem is that she is unjustifiably inferring lasagna-law from lasagna-fact.

The analogy to constitutivism should be obvious. The constitutivist suggests that we already have a reason to choose the moral law as our meta-principle, because the moral law (autonomy-law) is supposed to be equivalent to or entailed by our capacity to act intentionally (autonomy-fact). Since we are in fact autonomous, it follows that we always already give ourselves the moral law. But, like Megan's lasagna-constitutivism, metaethical constitutivism supposes that full-stop normative necessity follows from the fact that we are inescapably involved in some activity with some constitutive aim. And, barring further support, it does not, because it does not follow from the fact that we are condemned to exercise a certain capacity that we are condemned to *care* about the capacity itself.¹² This is the shmagent objection, which we can now characterize as the objection that autonomy-fact does not entail autonomy-law.¹³

The no-metaethics objection arises next. To see this, suppose that Megan decides to press me further. How can she do this? It seems that the only thing she can do is claim that the

¹² I think that Gibbard (1999, 154-155) addresses a similar point to an earlier version of Korsgaard's views. Gibbard's point is that valuing some *x* does not entail that one is required to value one's capacity to value in general.

¹³ Some constitutivists will want to say here that any putative skeptic belies his position as dialectically unstable, because just in contemplating whether or not to care about being an agent, he exhibits that in fact he does care (cf. Velleman, 2009; 142ff). This idea involves a combination of the familiar idea that agency is a game that you can't quit playing, coupled with the idea that you have to care about it to count as playing it. This issue is complicated, and I don't have room to worry about it here. Instead, I will quickly rehearse two responses to it, both due to Enoch (forthcoming). First, it does not seem that anyone who wants to *refute* the skeptic can accept a position like this – at best, it seems, on this view, you can deny that the skeptic really exists. Second, even if it turns out that we all do care about the constitutive norms of action, this is not uncontroversially relevant to the question, because the kind of necessity invoked in the response is metaphysical rather than normative. That we all do care is not sufficient evidence for the claim that we all should.

metaphysical fact that some activity is inescapable for me has a special normative status. But then, of course, I can wonder why. Even supposing that I am comfortable with the idea that some fact resolves the is-ought gap, I am entitled to an explanation of why that particular metaphysical fact is the one that does. And for all that the constitutivist account says, that explanation could be of any metaethical variety, including a Moorean non-reductive realist one.¹⁴ So the no-metaethics object can be characterized as suggesting that *if* autonomy-fact entails autonomy-law, Kantians have not told us *why*.

One way to read both of these objections, then, is as objections to the inference from autonomy-fact to autonomy-law. The claims, respectively, are that this is a bad inference (the shmagent objection) and that it is an unsupported one (the no-metaethics objection). What I have argued here is that, if these worries obtain, there is reason to think that they are derivative from the Kantian conception of autonomy, on which autonomy is supposed to be a categorical norm just in virtue of its also being a metaphysical fact. The suggestion that I am advocating is that if the inference is bad, it is bad because the metaphysical-fact aspect of the conception, the one on which the first premise of the constitutivist inference relies, is bad. If autonomy is a constitutive norm of an inescapable human activity, it had better not be a norm *because* that activity is inescapable. A different kind of story seems to be required.¹⁵

Georgia State University
P.O. Box 4089 / Atlanta, GA 30302-4089
ptulipana1@gsu.edu

References

- Buss, Sarah, 1999. "What Practical Reasoning Must Be if We Act for our Own Reasons", *Australian Journal of Philosophy*, 77, 399-421.
- Davidson, Donald, 1978. "Intending", reprinted in *Essays on Actions and Events* (Oxford: Oxford UP), 83-102.
- Enoch, David, 2006. "Agency, Shmagency: 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: Oxford UP), 303-333.
- Gibbard, Allan, 1999. "Morality As Consistency in Living: Korsgaard's Kantian Lectures", *Ethics: An International Journal of Social, Political, and Legal Philosophy*, 110:1, 140-164.
- Hussain, Nadeem and Nishi Shah, 2006. "Misunderstanding Metaethics: Korsgaard's Rejection

¹⁴ Kant (1997, *Ak.* 531) seems to want to exclude at least some varieties of realism, cautioning that this "strange" fact "is not an empirical fact but the sole fact of pure reason which, by it, announces itself as originally lawgiving".

¹⁵ Thanks to Christie Hartley and Hunter Thomsen for their helpful comments on this paper.

- of Realism", in Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics, vol. 1* (Oxford: Oxford UP), 265-294.
- Kant, Immanuel, 1997. *Critique of Practical Reason*, trans. Mary Gregor, (Cambridge, UK: Cambridge UP).
- 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).
- , 1999. "Self-Constitution in the Ethics of Plato and Kant", reprinted in *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (Oxford: Oxford UP), 100-128.
- , 2003. "Realism and Constructivism in Twentieth-Century Moral Philosophy", reprinted in *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (Oxford: Oxford UP), 302-326.
- , 2004. "Fellow Creatures: Kantian Ethics and Our Duties to Animals", *The Tanner Lecture on Human Values*.
- , 2008. "Introduction", in *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (Oxford: Oxford UP), 1-23.
- , 2009. *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford UP).
- Milgram, Elijah, 2009. "Practical Reason and the Structure of Actions", *The Stanford Encyclopedia of Philosophy (Summer 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2009/entries/practical-reason-action/>.
- Pinkard, Terry, 2002. *German Philosophy 1760-1860: The Legacy of Idealism* (Cambridge: Cambridge University Press).
- Railton, Peter. 1997. "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action", in Garrett Cullity and Berys Gaut (eds.), *Ethics and Practical Reason*, (Oxford: Clarendon), 53-79.
- Street, Sharon, 2006. "Constructivism about Reasons", in Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics, vol. 3* (Oxford: Oxford UP), 207-245.
- Velleman, J. David, 1996. "The Possibility of Practical Reason", reprinted in *The Possibility of Practical Reason* (Oxford: Clarendon), 170-199.
- , 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.