# Meeting constitutivists halfway

**Citation for published version:**
Ridge, M 2018, 'Meeting constitutivists halfway', *Philosophical Studies*, vol. 175, no. 12, pp. 2951-2968.
https://doi.org/10.1007/s11098-017-0989-z

**Digital Object Identifier (DOI):**
10.1007/s11098-017-0989-z

**Link:**
Link to publication record in Edinburgh Research Explorer

**Document Version:**
Publisher's PDF, also known as Version of record

**Published In:**
Philosophical Studies

Philos Stud
DOI 10.1007/s11098-017-0989-z

# Meeting constitutivists halfway

Michael Ridge[1]

**Abstract** Constitutivism is best understood as a strategy for meeting a set of related metanormative challenges, rather than a fully comprehensive metanormative theory in its own right, or so many have plausibly argued. Whether this strategy succeeds may depend, in part, on which broader metanormative theory it is combined with. In this paper I argue that combining constitutivism with expressivism somewhat surprisingly provides constitutivists with their best chances for success, and that this combination of views has some surprising benefits for both parties.

**Keywords** Constitutivism · Expressivism · Internalism · Meta-ethics · Metaethics

Constitutivism is best understood as a *strategy* for meeting a set of related metanormative challenges, rather than a fully comprehensive metanormative theory in its own right, or so many have plausibly argued (Hussain and Shah 2006; Silverstein 2012; Ridge 2012). Whether this strategy succeeds may depend, in part, on which broader metanormative theory it is combined with. In this paper I argue that combining constitutivism with expressivism somewhat surprisingly provides constitutivists with their best chances for success.

At the same time, this combination of views should help expressivists deal with some residual but often unnoticed problems. In particular, a synthesis of constitutivism and expressivism could help expressivists deal with the worry that it cannot adequately explain the universality of specifically *moral* reasons insofar as these ground obligations. The worry for expressivists is that moral obligations entail aptness for blameworthiness for failing to fulfil those obligations, but

✉ Michael Ridge
mridge@staffmail.ed.ac.uk

[1] University of Edinburgh, Edinburgh, UK

Published online: 14 October 2017

blameworthiness in turn presupposes that the relevant moral reasons must be *accessible* to the agent in a way that might seem difficult to explain within a broadly expressivist framework. It would be inappropriate to blame someone for failing ot act on reasons whose force they simply could not appreciate. Constitutivism can, I shall argue, help expressivists deal with this problem. Constitutivism and expressivism are, in these ways, made for each other.

However, to enter into this union, constitutivists must abandon some of their theoretical aspirations. Most obviously, they must concede that constitutivism is neither a fully free-standing meta-normative theory nor a way of transcending traditional metanormative debates. At least some constitutivists have been reluctant to concede this.

More interestingly, constitutivists must *also* abandon the idea that what it is for some consideration to be a reason for action just *is* for its status as a reason to somehow to be derivable from something which is constitutive of agency. Indeed, they must also abandon the thesis that necessarily something is a reason for action *only if* it somehow can be derived from something constitutive of agency. Instead they should hold that something is a reason for action *if* it can somehow be derived from something constitutive of agency. They must, in other words, go from an "if and only if" connecting something's being a reason for action from its being derivable from what is constitutive of agency to a mere "if". The constitutivist must, in this context allow for at least the possibility of reasons which *cannot* be derived from what is constitutive of agency. It is in this sense that expressivists and constitutivists can meet "halfway".

Ultimately, the arguments of this paper are entirely conditional. For the success of constitutivism about specifically moral reasons obviously depends on whether what is constitutive of agency can deliver anything with recognizably substantive moral content. That, of course, is a huge issue in its own right, and one I shall not be able to even begin to explore here. Instead, I shall simply argue that *if* the constitutivist can "deliver the goods" by showing how some feature constitutive of agency brings in its wake a commitment to a norm or aim with recognizably substantive moral content, then constitutivism and expressivism are "better together".

## 1 Constitutivism and its attractions

Constitutivism is usefully understood as a strategy for meeting sceptical challenges to the authority of some norm or end. Although constitutivism is available in other domains (most notably, the epistemic), in this paper I focus entirely on constitutivism in the practical realm, and uses of 'constitutivism' should henceforth be read as shorthand for 'constitutivism in the practical realm' unless otherwise indicated.

Suppose someone is sceptical about the authority of morality. This might simply be because they do not see why moral norms should be presumed to have rational authority. Alternatively, it might be because they have been impressed by some specific sceptical argument, like the argument Thrasymachus gives in book one of

*The Republic*, say. Whatever the source of the sceptical challenge, the constitutivist strategy for meeting it is the same.

The constitutivist tries to show how *simply being an agent* commits one to some norm or end with recognizably substantive moral content. The relevant norm or end is in this sense *constitutive* of agency. In that case, the sceptic is also committed to the relevant end or norm. The idea is that this makes sceptical challenges somehow self-stultifying.[1]

Kant and neo-Kantians are, of course, the most famous practitioners of this strategy. Kant can be read as arguing that the categorical imperative is constitutive of agency (Cf. especially Korsgaard 2008 and Korsgaard 2009). Kant himself tried to derive a commitment to the categorical imperative from the very idea of freedom of the will. Simply in virtue of being agents who deliberate about what to do, we must take ourselves to have freedom in a negative sense—roughly, in that our actions are not determined by anything external or alien. Free will in this sense, though, commits us to freedom in a positive sense—roughly, being self-directed or autonomous. Kant then argued that to be free in this sense is not to be lawless or random, but to be governed by a law which one must give to oneself simply qua rational agent, and he argued that this law is in fact the moral law which for us (unlike a "Holy Will") takes the form of a categorical imperative (see Kant 1998).

One nice thing about this Kantian version of constitutivism is its emphasis on autonomy. This allowed Kant to argue that all other attempted foundations of morality fail because they appeal to something external to the agent, thus making morality *heteronomous*. This line of thought illustrates why Kantian constitutivism is so attractive as a strategy for meeting sceptical challenges to the rational authority of morality. It suggests that *other* attempts to meet such challenges are doomed to failure because they appeal to norms which are external to the agent, and which can therefore themselves be intelligibly challenged. Indeed, on Kant's view, adherence such external norms as one's most fundamental practical stance would deprive one of autonomy. Moreover, because moral norms are on this account derived from agency as such, morality's pretensions to *universal* validity are upheld.

Another virtue of this strategy for meeting sceptical challenges to the authority of morality is that it can seamlessly explain how moral judgments can motivate us to act. At least, if we understand the commitment to a norm like the categorical imperative as a practical stance, then it is not hard to see why such a commitment could motivate someone to do the right thing. This stands in stark contrast to some forms of so-called "externalist cognitivism" according to which moral judgments are simply representations of external facts of some kind. Whether the relevant facts

[1] A different but closely related strategy is to appeal not to what is constitutive of agency as such, but to what is constitutive of *ideal* agency. This is the strategy pursued by Michael Smith in e.g. Smith (2015). For useful critical discussion of Smith's approach, see Bukoski (2016). I lack the space here to engage with the nuances of Smith's distinctive approach and its relationship to more standard constitutivist appeals to what is constitutive of agency as such. I do, however, discuss Smith's idealized subjectivist analysis of reasons for action in the following section, not to engage with his conception of constitutivism, but simply to explore one way of filling a lacuna in the constitutivist programme. Thanks to an anonymous referee for pressing me on this important difference between Smith's constitutivism and other species of the genus.

are purely naturalistic or are instead facts about non-natural normative properties, it is at least not obvious why the recognition of such facts could in itself motivate someone to do the right thing.[2]

Finally, constitutivism *might* be able to meet sceptical challenges to the authority of morality without any dubious metaphysical or epistemological commitments. Ultimately, this will depend on the details of the constitutivist's theory of agency. We need look no further than Kant himself, who located rational agency in a purely "noumenal" realm, to see how constitutivists might have to undertake controversial metaphysical and epistemological commitments. However, it is not at all obvious that these commitments really are *essential* to the constitutivist project—even to a recognizably Kantian version of it. At least some modern day Kantians try to show how the core insights of Kant's practical philosophy can be preserved in a more broadly naturalistic worldview that does not require Kant's transcendental idealism.

Of course, it is at best deeply controversial whether Kant or neo-Kantians *can* derive the categorical imperative (or *any* recognizably moral norm) from rational agency as such.[3] Nor is the Kantian approach the only game in town. Humean and Nietzschean forms of constitutivism have also been developed in some detail, though whether these would, even if successful in their own terms, vindicate recognizably moral norms is very much up for debate.[4] I am here going to make the rather large assumption for the sake of argument that some version of constitutivism can succeed at least in showing how a norm or end with recognizably substantive moral content can be derived from features constitutive of agency itself. Even granting this much, the constitutivist is by no means home free, and it is instructive to see why this is the case and how the constitutivist programme might best be filled out. In the next section I explain how constitutivism must be paired with some broader metanormative theory before we can properly assess its tenability.

## 2 Some problems for constitutivism

The idea that some norm or end is constitutive of agency can be understood in at least two ways, and constitutivists are not always as clear as they might be about this. First, a norm or end might be constitutive of agency in the sense that agency presupposes a belief that the relevant norm or end is good, rationally compelling, warranted or has some other normative property. 'Belief' here should be understood as a state of the same kind as ordinary descriptive beliefs, but with a normative content; the intended contrast is with 'belief' in the broader sense in which quasi-realist expressivists allow that normative judgments are beliefs. Call this the

[2] See Katsafanas (2013): chapter one for a more extensive discussion of the issue of practicality and constitutivism. Katsafanas emphasizes that morality not only can motivate us but that it can also seem to *constrain us*, and he argues that constitutivism (at least, the sort of constitutivism he favours) can help make sense of this sense of constraint as well. I lack the space to delve into this important nuance here.

[3] For an important recent attempt to vindicate Kant, see Korsgaard (1996). For some critical discussion of Korsaard's attempt to vindicate the Kantian argument, see Ridge (2005).

[4] See Street (2008) for a Humean form of constitutivism and Katsafanas (2013) for a Nietzschean one.

"cognitivist" reading of constitutivism. Second, a norm or end might be constitutive of agency in the sense that agency presupposes a *practical* commitment to that end or norm—willing the end or accepting the norm in something like Allan Gibbard's sense (see Gibbard 1990).[5] Call this the "practical" reading of constitutivism.

As a number of commentators have pointed out, the cognitivist reading raises some vexatious questions for the constitutivist.[6] First, what exactly is the content of these normative beliefs? If the content is naturalistic, then the constitutivist view collapses into a form of naturalism, whereas if the content is not naturalistic then it presumably collapses into a form of non-naturalism.[7] Either way, constitutivism collapses into another familiar metanormative view.

This need not in itself be embarrassing for the constitutivist, though constitutivists who insist that their view is an entirely new and fully free-standing metanormative theory will, of course, be disappointed. It might, though, be that constitutivism is best understood as a distinctive *form* traditional metanormative views can take, and which provides unique advantages. In fact, I think this is the most fruitful way to think about constitutivism.[8]

A much more troubling question, though, is what grounds we have for thinking the relevant belief is *correct*. This is a standard worry about transcendental arguments. Granted, we cannot help but think that p insofar as we are agents—how does this show that p is actually *true*? Insofar as we think of these beliefs as representing a way the world might be, it is hard to see why the fact that we cannot but think that the world is that way indicates that it really is.[9] This sort of worry is familiar in other contexts. Consider the case of free will. Perhaps, as some have argued, as agents who deliberate about what to do we cannot but take ourselves to have a robust form of free will. That in no way shows that we really *do* have free will. The world might simply not cooperate.

More could be said about this approach, but this should already be enough to explore the practical reading. Moreover, the practical reading fits better with much of what constitutivists say about their view than the cognitivist reading. Korsgaard, for example, heaps scorn on what she calls "substantive realism" and seems to understand our commitment to the categorical imperative as a purely practical matter that arises in first-person deliberation (Korsgaard 1996). Paul Katsafanas is quite explicit that his constitutivism should be understood in terms of a practical commitment to a certain set of ends (Katsafanas 2013). Sharon Street carefully distinguishes normative judgments (in one sense, anyway) from ordinary beliefs and emphasizes their practical and desire-like role (Street 2008; see also Ridge 2012).

[5] See also Ridge (2014: chapter 4).

[6] See, e.g. Hussain and Shah (2006), Ridge (2015) and Silverstein (2012).

[7] Another possibility is that normative judgments have their content not in virtue of what they represent, but in virtue of their inferential role. In this case, the view again seems to collapse into another view, namely metanormative inferentialism. See Chrisman (2008) and Chrisman (2016). Compare Wedgwood (2007) for a view which tries to derive a realist form of non-naturalism from a broadly inferentialist or "conceptual role" approach to semantics.

[8] See Ridge (2012).

[9] This point is not a novel one. It is well made in Silverstein (2012: p. 8, e.g.), for example.

Suppose it could be shown that willing a recognizably moral end (like treating humanity with respect) was partly constitutive of being a rational agent. By itself this does not obviously entail that one has good reason to promote that end. Two strategies suggest themselves for bridging the apparent gap between the thesis that willing some end is constitutive of being an agent and one's having good reason to promote that end. The first, which I will discuss in the remainder of this section, is to adopt some form of subjectivist cognitivism. The second, which I shall argue (in Sect. 3) is more promising, is to combine the constitutivist thesis with a quasi-realist form of expressivism.

The simplest way to derive reasons from constitutive ends is to appeal to a simple subjectivist theory according to which having a reason to perform some action just *is* for that action to promote and end you will. One virtue of this simple subjectivist account is that it explains how having a justifying reason can *explain* one's acting on that reason. The intuitive appeal of preserving some such link between what are sometimes called motivating and normative reasons was famously one of the main attractions of Bernard Williams' famous suggestion that we understand justificatory reasons as "internal" reasons, where being internal is a matter of being suitably connected to an agent's "motivational set".

Moreover, a simple subjectivist conception of reasons might seem to help ensure that agents are not deeply *alienated* from what they have reason to do. Once again, this idea can be found in Bernard Williams' discussion. Williams uses the memorable example of Owen Wingrave's supposedly having an "external reason" to fight in World War I to highlight why one might be worried about the very idea of external reasons (Williams 1979; see also Schroeder 2007 for a more recent attempt to defend a broadly subjectivist conception).

It is not hard to find the appeal to some such subjectivist conception in the work of many constitutivists. Paul Katsafanas, for example, is very clear that for any X (whether an agent or not), insofar as X aims at G, G is a standard of success for X (Katsafanas 2013: p. 39). From this "relatively uncontroversial" claim, he infers that we can derive justificatory reasons from agents' aims (Katsafanas 2013: p. 14). So whenever an agent aims at something, this provides the agent with reasons to do whatever would promote the object of that aim. All that is special about constitutive aims is that they are universal and cannot be shed in light of conflicting with some other aim. Sharon Street also is plausibly read as appealing to a subjectivist conception of reasons for action (for an extensive argument for this reading, see Ridge 2012).

Although this simple subjectivist strategy has some independent attractions, it also has some dramatically counterintuitive results. Simply focus on someone whose aims are banal, self-destructive or evil to see how the proposed account has implications we might like to avoid. John Rawls' memorable example of someone whose ambition is to count blades of grass, or Allan Gibbard's "maximally coherent anorexic" or a ruthless Mafioso all have ends we might like to be able to hold provide no reason whatsoever.

These examples are all cases in which the theory seems to predict "too many reasons"—reasons we intuitively do not think really are reasons. The account also in some cases seems to provide too few reasons. For example, the account also implies that someone who does *not* care about her own welfare at all (e.g.) thereby has no reason to look after herself. Rather, she will have no such reasons apart from the ways in which looking after herself would promote some other end she cherishes. However, this link will seem far too contingent and also the wrong sort of reason to most people.

Obviously a lot more can and has been said about this, but to my mind this is already enough to at least motivate the consideration of alternatives.[10] The strategy is also, of course, open to standard worries about any reductionist form of cognitivism. For example, Moore's "Open Question Argument" and the Horgan and Timmons' "Moral Twin Earth" style arguments are relevant here. Which of these is salient will depend on the details of the proposed subjectivist theory—analytic forms of subjectivism must face the Open Question Argument, while a posteriori forms of subjectivism (on the model of water is H20 a la Putam) must face the Moral Twin Earth challenge.

Putting these standard objections to any form of reductionist cognitivism to one side, the worry that a simple subjectivism entails too many reasons—reasons to do things that are inane or evil just because the agent happens to desire it—is a powerful one. Perhaps, though, we can adopt a more sophisticated subjectivism to avoid this worry. One strategy would be to endorse an idealized form of subjectivism, according to which our reasons are fixed not by what we contingently happen to desire, but by what we would desire if idealized in various ways—if our desires were more coherent and fully informed, say. Crucially, any desires which are constitutive of agency as such be desires had by any idealized agents, simply in virtue of their still being *agents*.

The most famous contemporary defender of this sort of idealized subjectivism is Michael Smith (Smith 1994). Smith rightly emphasizes that what any given non-ideal agent has reason to do should be fixed not by what his ideal self would *do* if in his circumstances, but by what his ideal self would want his non-ideal self to do in his non-ideal circumstances—warts and all, as it were. After all, my ideal self will not have reason to do various things simply in virtue of being ideal that I, being very far from ideal will have strong reason to do—e.g. get more information and reflect more on the coherence of my desires.

One problem with simple subjectivism was that it generated too many reasons. Idealized subjectivism can certainly mitigate this problem to some degree. In many instances, people desire the banal or the evil because they have not reflected enough on how these desires cohere with their other desires, or the desire rests on ignorance of relevant facts or on outright false beliefs. However, the "too many reasons" objection is plausibly not entirely met in this way. On a standard way of understanding the relevant idealizations, they are all *procedural* -they involve gathering more facts, correcting false beliefs, and making one's overall set of desires more coherent. In principle, there is nothing to prevent this process from culminating in a fully coherent set of banal or evil desires. "Garbage-in/Garbage-out" is an apt slogan here. Allan Gibbard gives the memorable example of the "fully coherent anorexic" (Gibbard 1990: p. 166).

[10] Again, see Schroeder (2007) for a recent attempt to defend a desire-based view. Schroeder discusses both the problem of "too many reasons" and the problem of "too few reasons" at length.

Smith's own analysis arguably sidesteps this worry. This is because Smith's analysis of reasons for action requires that there be *convergence* in the desires of our idealized selves—indeed, of the idealized versions of all *possible* rational agents. Insofar as we have reasons for action at all only insofar as the desires of our idealized selves converge, it cannot turn out that the enormous variety of desires with which we might start out can matter to our reasons. Somehow these contingent desires must be *filtered out* by the relevant idealization.

I do not have the space here to engage with the details of Smith's own interesting attempts to derive a plausible set of reasons from this approach. At the outset, though, it does seem unlikely that there will be convergence on *any* substantive desires. After all, the process of idealization involves only making our desires more informed and more coherent. Insofar as we start out with radically different desires, it seems very plausible that making those desires more informed and coherent will not lead us to converge. Once we widen the scope of our analysis to include all *possible* rational agents, the idea that there would be any substantial convergence becomes, prima facie anyway, highly problematic.

Idealized forms of subjectivism, like Smith's, are more promising when it comes to not delivering too many reasons. However, these views instead threaten to deliver *too few* reasons—possibly no reasons whatsoever, in fact, if there is no convergence of the needed kind. Smith himself argues that our ideal selves would converge on a desire that our actual selves desire to not interfere with their current exercise to know the world in which they live, a desire not to interfere with their current exercise of their capacity realize their intrinsic desires, a desire not to interfere with their future selves' efforts at desire satisfaction, an desire to develop their capacities to know the world and promote their desires, and a desire to promote these capacities and their exercise in others (Smith 2015: p. 191).

The argument that our ideal selves would have these desires is in my view problematic, but I lack the space to engage with Smith's subtle arguments here (see, though, Bukoski 2016 for useful discussion). However, I think the arguments are controversial enough that insofar as the tenability of constitutivism is dependent on their soundness we would do well to look for other ways to preserve whatever we find insightful in the constitutivist programme. More to the point, Smith's argument that our ideal selves would have these substantive desires does *not* appeal to the theses that these desires are constitutive of agency as such, and so trivially would be constitutive of ideal agency. Rather, he appeals to the idea that agency is a "goodness fixing kind" and that these desires can be derived instead from the function of agency (though in what sense of 'function' is not entirely clear). To that extent, Smith's strategy is not a form of constitutivism in the sense I am exploring in this paper.

Of course, a constitutivist of the kind I have in mind could combine their constitutivism with Smith's analysis of reasons for action. They would then have the burden of demonstrating that *all* our reasons for action could be derived from features constitutive of agency as such. This is, to say the least, a tall order. Even restricting our attention to the moral sphere, existing constitutivist arguments seem likely to deliver at best what T.M. Scanlon usefully called the "morality of what we owe to one another" (Scanlon 2000). Kantian constitutivist like Korsgaard can, at best, deliver the kind of morality Kant was after—a morality exhausted by the idea of mutual respect between free and equal rational agents. To his credit, Scanlon himself emphasizes that this does not exhaust common sense morality. Contra Kant, many ordinary folks think we have duties to sentient nonhuman animals and to the environment. Scanlon also points out that certain aspects of sexual morality are unlikely to be well analysed in Kantian terms. Of course, one might hold out hope that a *further* constitutivist argument could show that rational agents as such have commitments e.g. to care about the natural environment for its own sake, to be kind to sentient but non-rational agents, etc. In my view this even more ambitious constitutivism is ex ante very unlikely to succeed, and it is no coincidence that Kantians do not typically try to derive such further norms from agency as such. Moreover, this is still just restricting our attention to the moral sphere, broadly conceived. A fully convincing constitutivism of this sort would also need to derive norms or ends which could explain and justify our prudential reasons, our aesthetic reasons, and so on.

It would be better, then, if we could somehow accommodate what seems insightful and important in the constitutive strategy without committing ourselves to a simple subjectivism which entails too many reasons, or instead adopting an idealized subjectivism like Smith's which likely entails too few. Fortunately, there is a way to do this. By combining constitutivism with quasi-realist expressivism, we can meet the constitutivist halfway in the following sense: We can agree with the constitutivist that *if* there are ends or norms which are constitutive of agency, then those ends or norms do indeed provide universal reasons for action. However, we need *not* agree with the constitutivist that our reasons for action are *limited* to whatever can be derived from those ends or norms constitutive of agency. Insofar as the combination of expressivism with constitutivism provides a principled basis for meeting the constitutivist halfway in precisely this way, it is a very attractive package, or so I shall now argue.

## 3 Meeting the constitutivist halfway

I am not the first to note that constitutivism and quasi-realist expressivism might fit well together. Jay Wallace, for example, notes that because constitutivism is addressed to a different set of questions from expressivism, the two do not seem to be in direct competition (though he uses the term 'constructivism' it is clear he has in mind forms of constructivism which are also constitutivist in my sense):

> Thus, for all I can see, expressivism and constructivism may be compatible with each other. The constructivist could endorse the expressivist explanation of the semantic and logical properties of normative discourse…and it is open to expressivists for their part to accept what the constructivist says about the nature of normativity (Wallace 2013: pp. 26–27).[11]

[11] Compare Lenman (2013).

However, Matthew Silverstein has provided the most extensive discussion to date of why expressivism and constructivism are not only compatible but fit very well together. Here I want to amplify and build on his arguments, so I begin simply by summarizing his main points.

Silverstein usefully highlights the impressive extent to which prominent constitutivists use arguments and even rhetoric that comes straight out of the expressivist play book. David Velleman, for example, suggests that we forgo transposing demands into indicative judgments but instead leave them in their practical form as demands (Velleman 2009: p. 116). To take another example, Korsgaard is up to her neck in expressivist-sounding arguments and ideas. She objects strenuously to substantive moral realism which she sees as dogmatic foot-stamping which refuses to engage with the hard problems of practical philosophy in a convincing way. She emphasizes that normative questions are not third person or theoretical questions about how the world is, but are instead “first person” practical questions which arise only when deliberating about what to do. Normative concepts function to solve practical problems. She at one point allows that both realism and expressivism are true “in their way,” but in that context by realism she seems primarily to mean that normative questions have objectively correct answers. A quasi-realist expressivist would not deny this, of course, so it is not clear that the realism which is “true in its way” is actually a rival to quasi-realist expressivism. On the same page, Korsgaard goes on to assert that expressivism is true but “in a way that makes it boring”. (Korsgaard 2008: p. 325, n. 490).

Suppose that the constitutivist takes what in the previous section I called the “practical” approach. In that case, there will be practical states (either the adoption of ends or the acceptance of norms) which are constitutive of agency. *If* these practical states are of the same type that the expressivist theory classifies as *being* normative judgments, then there will be normative judgments we cannot help but make simply in virtue of being agents.

This might seem dangerously close to the cognitivist interpretation discussed in the previous section, since we can now wonder whether these judgments are correct. However, as Silverstein points out, the quasi-realist conception of these judgments limits the kinds of doubt we can entertain about the correctness of normative judgments to doubts which arise *internally*—that is, within the scope of our engaged normative deliberations. The idea that we might somehow stand outside our normative commitments and worry about whether they correspond to some putative normative reality which we can conceptualize independently of making any specific normative judgments is entirely alien to the quasi-realist approach. While quasi-realists do quite rightly insist on a theory of error, the standard approach is to focus on how our judgments might change if we were improved in various ways, but where what counts as an “improvement” is itself a normative question.[12]

Silverstein is right; by “going expressivist,” the constitutivist can avoid one of the main problems canvassed in the previous section. However, Silverstein’s

[12] See Blackburn (1996). A worry about this approach is whether it can make sense of what Egan calls “fundamental moral error”—see Egan (2007). For Blackburn’s reply to Egan, see Blackburn (2009). For further discussion see Kohler (2015) and Ridge (2015).

characterization of the details of this move are problematic. He argues that *all* that matters to the expressivist vindication of the relevant norm or end is "the value to which we are inescapably committed, *not the fact that we are inescapably committed to it*" (Silverstein 2012: p. 13, emphasis added). His reason for this is that while normative inquiry occurs entirely from within the practical point of view, given expressivism, the inescapability is "something we observe from outside the practical point of view" (Silverstein 2012: p. 13).

In my view, this rests on an impoverished conception of the practical point of view. The crucial point is that the ends which are constitutive of our agency might *conflict* with some of our contingent ends. For example, a CEO's end of maximizing profits conflicts with his also treating humanity with respect, which (let us assume) is an end which is constitutive of his agency. Plausibly, when two ends conflict we are *rationally required* to abandon at least one of those ends.

However, it is also plausible that the rational 'ought' implies 'can', and that this constraint on our deliberations is one that quite correctly plays a role in our practical deliberation. In that case, though, *the fact that an end is inescapable can and should appear from within the practical point of view*. Just as the fact that I cannot swim the English channel on my own should appear on my practical radar if I am deliberating about whether to make that my end, the fact that I cannot (qua agent) abandon my end of treating humanity with respect should appear on my practical radar when deliberating about whether to abandon that end for the sake of some other end.[13]

The idea of inescapability must be handled with care, and must be understood in a way that is compatible with the kind of Kantian constitutivism that I think provides the most promising fit with expressivism. In particular, clinical conditions we might sometimes call "psychological incapacities" should not count as incapacities in the relevant sense.[14] Insofar as a kleptomaniac remains a rational agent at all, and hence acts under the idea of freedom, there is an important sense in which she can abandon the end of stealing. The Kantian line on such cases is to distinguish mere inclinations from maxims. The kleptomaniac may feel an incredibly strong inclination to steal, and this will make it very difficult for her to refrain from stealing. Insofar as she remains a rational agent at all, though, her inclinations cannot be taken (from the first person point of view) to determine her will. It is still up to her whether she incorporates the object of this inclination into a maxim. If we stipulate that it is impossible for her to do this, then when it comes to this issue we cease to treat her as an agent—in which case it will not be true that she ought not steal, but only because she is not subject to norms at all once it is clear that has no agency in the needed sense with respect to this decision. The contrast, of course, is with an incapacity which stems from one's status as a rational agent.

The broader point is that, properly understood, the inescapability of the relevant ends (or norms) can and should play a role when an agent is tempted by some contingent end with which it conflicts. In fact, this is another virtue of the blending of expressivism and constitutivism—it helps the expressivist accommodate the idea

[13] Compare Katsfanans (2013: pp. 187–188).

[14] Thanks to an anonymous referee for drawing me out on this point.

that moral duties *trump* other kinds of reasons—insofar as non-moral reasons are not also fixed by ends which are constitutive of us as agents.[15]

To make this more concrete, suppose (very optimistically) that it could be shown that treating humanity with respect is an end which is constitutive of agency. Suppose, moreover, that for the sake of making a financial gain, I am tempted to lie to someone in circumstances in which I judge that lying to them in this way would be failing to treat their humanity with respect. Suppose, finally, that not only is treating humanity with respect a constitutive feature of agency, but that I have been convinced of this. I can then reason as follows: My ends conflict—I cannot both gain this financial end *and* treat humanity with respect. I can know a priori that when I have conflicting ends that I must give up at least one of them on pain of irrationality. This means that I must give up one of these two ends to avoid irrationality. Since, it seems, I *cannot* give up the end of treating humanity with respect (Kant convinced me of this), I am left with only one option insofar as I want to remain rational—giving up the end of financial gain in this circumstance. I shall therefore abandon the end of making that financial gain and refrain from lying to this person. Crucially, the fact that the end of treating humanity with respect is *inescapable* played a vital role in this line of reasoning.

One might worry that vindicating the hypothesis that morality trumps other reasons is a cost rather than an advantage of the view.[16] Intuitively, one might have thought, sometimes one's own self-interest should come ahead of moral considerations. For example, if I would have to break a promise to a friend to meet her for dinner to make it to an important job interview, then it might well make most sense to break the promise. To see why this seductive worry goes wrong, we must first distinguish the thesis that moral reasons trump other reasons from the thesis that moral *duties* trump other reasons. To see this we must first take a brief diversion into the kind of Kantian framework a successful constitutivism might hope to deliver.

Suppose constitutivism delivered a standard Kantian moral theory. It is famously hard to find mere pro tanto moral reasons, as opposed to moral duties, within Kant's framework. Hard, but perhaps not impossible. Kant's conception of imperfect duties can be interpreted in a way that delivers such reasons. On the sort of interpretation I have in mind, the imperfect duty of beneficence can be understood as a kind of perfect duty to always count the fact that one's action would promote another rational agent's permissible ends as *a reason*, but perhaps not a decisive reason, to perform that action. The categorical imperative itself does not entail how much weight to assign such reasons; that determination is left with the agent. Although an agent could in principle count such considerations as reasons but never help anyone

[15] Actually, it will only be those moral duties which can be derived from what is constitutive of agency that automatically rationally trump other reasons. Insofar as I am allowing that e.g. our duties to nonhuman animals might be moral but not grounded in the constitutivist way, these reasons will not automatically trump any conflicting non-moral reasons. It is not clear to me that this is on balance a worrying implication, though it will depend on how we understand the way that moral reasons trump non-moral reasons and also on the content of morality. I say more about this more general issue in the main body of the text.

[16] Thanks to an anonymous referee for pressing me on this point.

else, the more often an agent foregoes such opportunities, the more reasonably we may doubt her sincerity in counting the fact that an action would help someone as a reason.

Returning to the issue of morality trumping other kinds of norms, the first point is that mere pro tanto moral reasons, like the reasons generated by Kant's imperfect duty of beneficence on the interpretation I just sketched, need not carry the day at all, and so need not trump non-moral reasons. It is only proper moral *duties* that trump other kinds of reasons on the view developed here.

Obviously this is also controversial and a large topic in its own right which I cannot possibly hope to address properly here. In my view, though, the idea that moral duties are rationally overriding has some intuitive appeal. Moreover, it arguably makes better sense of the link between moral duty and blameworthiness. Plausibly, if someone fails to do their moral duty then, unless they have some excuse, they are blameworthy. Yet blaming someone for doing something they did not have most reason to do seems incoherent. The following speech-act seems highly problematic, after all: "In violating your moral duty, you did what you had most reason to do of course—your own self-interest in that case provided a weightier reason than the duty to tell the truth. Still, you are a real bastard for not doing your moral duty and you should feel very guilty about it!"[17]

One last point worth registering is that the intuitive plausibility of the thesis that moral duties override other kinds of reasons will ultimately depend in part on the content of those duties. If, for example, we were to agree with Kant that it is always morally wrong to tell a lying promise then the idea that moral duty overrides all other reasons will rightly seem far more counter-intuitive than if Kant's absolutism can in some way be tempered as some of his sympathetic commentators have tried to do. Here I shall simply have to make the optimistic assumption that the morality delivered by the constitutivist is sufficiently plausible that the thesis that moral duties override other reasons will not seem especially problematic.

I am here assuming not only that it is true that we endorse certain ends or norms simply in virtue of being rational agents, but that ordinary people know that they endorse these ends or norms. This, though, is fair enough since the *point* of constitutivism as a strategy is to provide a rationale which would be compelling to a deliberating agent insofar as the agent was convinced of constitutivism's core claims. I am here also assuming that people can be motivated in practical deliberation by their practical commitment to basic norms of practical rationality—in particular, a norm which forbids our simultaneously willing incompatible ends. Fortunately, I think we can and do reason in this way. Indeed, this is precisely what Kant had in mind when he famously remarked that while everything in nature acts in accordance with laws, only a rational agent can act *from* its conception of the law. We do this in the theoretical case when we reason with the explicit aim of avoiding contradictory beliefs too.

[17] This argument is not original to me. Stephen Darwall makes an argument very much like this in his *The Second-Person Standpoint*. See Darwall (2006). I point comes up again below.

This line of argument might seem in tension with what I consider an important insight from another of Silverstein's papers on constitutivism. In Silverstein (2015), he argues that it is not the inescapability as such of an end or norm which matters to the constitutivist project. This is for two reasons. First, the ends are not inescapable insofar as we can choose to renounce our status as agents—and sometimes we might well have good reason to do that. Second, though, it is not mere inescapability that explains why these ends or norms have the kind of authority the constitutivist aims to capture. After all, we do not in general think that inescapability confers authority.[18]

These points strike me as important and correct, but properly understood they are not in tension with the point made above about the relevance of inescapability to deliberation in certain contexts. My point there was *not* that inescapability conferred authority, but rather that inescapability could be relevant to deliberation simply because 'ought' implies 'can'.

Nor need I rely on the dubious idea that agency itself is inescapable. Granted we can and sometimes should abandon our status as agents. However, if we are deliberating as between which of two ends to abandon in our future agency, this option is not on the table, so that issue simply does not arise in the contexts I envisioned. In a more positive vein, I also agree with Silverstein that it is not the inescapability of the relevant norms/ends that matters, but our inability to somehow "stand outside" of these norms and ends and question them. It is not so much their inescapability but their unchallengability that matters.

Let us take stock. Combining constitutivist with quasi-realist expressivism avoids the problems canvassed for the constitutivist in the previous section. Unlike cognitivist interpretations of the commitments constitutivism posits, this combination avoids external worries about the correctness of normative judgments which are constitutive of our agency. Moreover, given the inescapability of the relevant norms/ends, it can at the same time explain why moral reasons trump other reasons—insofar as moral reasons can be given a constitutivist foundation, anyway.

Unlike the conjunction of subjectivism and constitutivism, it does not it posit to few or too many reasons. It also avoids Moorean "Open Question Argument" worries about such reductive naturalist interpretations and also avoids worries about the metaphysics and epistemology associated with non-naturalist realism. As a form of expressivism, the view accommodates the practicality of normative judgment while still allowing the constitutivist to defuse sceptical challenges to the authority of the moral reasons.

The benefits of this marriage are not so one-sided as this discussion might suggest. Not only does "going expressivist" help the constitutivist along these numerous dimensions, "going constitutivist" can help the expressivist too—or, rather, assuming as I have throughout that some recognizably moral norms or ends can be shown to be constitutive of agency.

In particular, the expressivist emphasis on the ubiquity of deep moral disagreement threatens to undermine an important aspect of our ordinary moral

[18] Contrast Fererro (2009).

practice. We routinely and confidently hold people morally responsible for their actions even when their own values support their actions and when it is not at all obvious that there would be any rational path from those values to what we consider to be the correct values. To be sure, sometimes there will be. Sometimes people's prejudices and moral vices are due to ignorance, false belief or some form of practical/normative incoherence. However, sometimes it might *seem* like the person's values are dubious in a way that could not be corrected in any of these ways.

The reason this is problematic is that it seems problematic to blame and punish people for doing something which not only was entirely rational given their values, but which they could not have rationally decided to refrain from doing. Granted, we might well be justified in punishing such people as a matter of self-defence, but pre-theoretically we often seem to think that such blame and punishment is fitting in a more broadly retributivist way—the culprits *deserve* our disapproval and sanctions. Yet there is something odd if not downright incoherent about blaming and punishing someone, telling them how horrible they are, but in the same breath allowing that they did what was rational, given their values at the time and that, moreover, there was *no* rational path from those values which could have led them rationally to choose otherwise.[19]

It is worth underscoring that this is especially a problem for expressivists. Granted, the intuitionist epistemology of moral realists is problematic in its own ways, but it is at least open to the moral realist to insist that the most fundamental moral truths are in principle open to anyone capable of clearly thinking about the issues who is also competent with moral concepts. Expressivists cannot appeal to our apprehension, however inchoate, of some normative reality which exists independently of us to make sense of how evil-doers in some sense "should have known better".

Constitutivism, if it works to vindicate moral norms or aims at all, could provide some obvious and much needed help here. Insofar as someone is an agent, they are *already* committed to the relevant moral norm or end. This should, in principle, give them the ability to derive their moral duties—so long as they are not blamelessly ignorant of some relevant descriptive fact, anyway. As far as I can see, there is no reason an expressivist who endorses a constitutivist story about moral norms cannot endorse this point, thus avoiding an otherwise serious worry about the extent to which we blame wrongdoers whose moral values we deem corrupt yet internally coherent. Given constitutivism about morality, even the most hardened and vicious criminals have it within themselves to derive their moral duties.

One might worry that one descriptive fact (in a broad sense of 'descriptive') that ordinary people might not appreciate is that the relevant ends or norms are *inescapable*. After all, one might have thought that you would need to be a professional philosopher who understands the ins and outs of constitutivism to appreciate the subtle fact that the value of humanity (e.g.) is somehow inescapable. Since the fact that the relevant end cannot be abandoned played a crucial role in the

[19] Compare Darwall (2006).

derivation of the rationality of doing one's duty, this looks like it could afford the wrongdoer with a sort of loophole.[20]

Here it is worth distinguishing a minimal version from an ambitious version of the constitutivist strategy I just sketched. The minimal version will help itself only to the idea that ordinary people will know that they endorse the relevant ends or norms. This assumption is fairly minimal since one might think that mere reflection and introspection would allow agents to know their own minds in this way. The idea would be that the fact that ordinary agents not only value humanity but know that they value humanity, this makes it more reasonable for us to blame them for treating humanity poorly *even if* they could not be expected to know that their valuing humanity cannot be abandoned. Simply valuing humanity and knowing that they do should give them the resources for rationally deciding to treat humanity with respect. We would, then, need some further account of why we could expect them to give this value more weight than a competing value, but here we might make the same kinds of common sense moves that we do in fact make when people try to justify bad behaviour. At any rate, some progress seems to have been made simply by establishing that the criminal had values (and knew she had values) which could have motivated her to do the right thing.

The more ambitious version of this strategy maintains that ordinary agents either know or could reasonably be expected to know that the relevant ends/norms are inescapable. This might seem incredible, since one might think to know this one would have to understand and accept the constitutivist theory which explains why the relevant end/norm is inescapable. This, though, rests on an unduly narrow conception of how one might come to discover that the relevant end/norm cannot be abandoned. One might instead learn this simply by *trying to abandon the end and finding oneself failing to do so.* Insofar as one can know one's own mind, one presumably will know that one still wills the end (e.g. of valuing humanity) even after one attempts to abandon it. Insofar as the constitutivist theory is right, then, one could discover this sort of volitional incapacity through trial and error. Nor is this sort of thought a million miles from ordinary moral thought and discourse. Ordinary people do characterize certain forms of behaviour as "unthinkable," and there is no obvious reason we should not take them at their word.

## 4 Conclusion

I have argued in this paper that constitutivism and expressivism are, in a way, made for each other. So long as the constitutivist is willing to give up on the overly ambitious thesis that *all* of our reasons for action must be derived from what is constitutive of agency, and instead hold that being derivable from what is constitutive of agency is instead merely *sufficient* for a reason for action, constitutivism combines very well with expressivism. If the constitutivist can in

[20] Thanks to an anonymous referee for pressing me on this point.

this sense meet the expressivist halfway then both parties can benefit substantially from the resulting union.

Of course, all of this simply *assumes* that the constitutivist can demonstrate that some recognizably substantive moral norms or ends can be shown to be constitutive of agency in the requisite sense. It is far from obvious to me that this can be done, and I have not yet seen an attempt which looks all that promising.[21] At the same time, though, it is not entirely obvious to me that this *cannot* be done, and it therefore seems to me worth figuring out why it might matter if it could be done. For now, those of us with expressivist sympathies should reserve judgment but wait to take advantage of any successful arguments for a recognizably moral constitutivism within an expressivist framework. Nice work if we can get it, but whether we can get it remains to be seen.

## References

Blackburn, S. (1996). Securing the nots. In: W. Sinnott-Armstrong (Ed.), *Moral knowledge?*. Oxford: Oxford University Press.

Blackburn, S. (2009). Truth and a priori possibility: Egan's charge against quasi-realism. *Australasian Journal of Philosophy, 87,* 201–213.

Bukoski, M. (2016). A critique of Smith's constitutivism. *Ethics, 127,* 116–146.

Chrisman, M. (2008). Expressivism, inferentialism, and saving the debate. *Philosophy and Phenomenological Research, 77,* 334–358.

Chrisman, M. (2016). *The meaning of 'ought'*. Oxford: Oxford University Press.

Darwall, S. (2006). *The second person standpoint*. Cambridge, MA: Harvard University Press.

Egan, A. (2007). Quasi-realism and fundamental moral error. *Australasian Journal of Philosophy, 85,* 205–219.

Ferrero, L. (2009). Constitutivism and the inescapability of agency. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 4, pp. 303–334). Oxford: Oxford University Press.

Gibbard, A. (1990). *Wise choices, apt feelings*. Cambridge, MA: Harvard University Press.

Hussain, N., & Shah, N. (2006). Misunderstanding metaethics. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 1, pp. 265–294). Oxford: Oxford University Press.

Kant, I. (1998). In M. Gregor (Ed.), *Groundwork of the metaphysics of morals*. New York: Cambridge University Press.

Katsafanas, P. (2013). *Agency and the foundations of ethics*. Oxford: Oxford University Press.

Kohler, S. (2015). What is the problem of fundamental error? *Australasian Journal of Philosophy, 93*, 161–165.

Korsgaard, K. (1996). *The sources of normativity*. Cambridge: Cambridge University Press.

Korsgaard, K. (2008). *The constitution of agency: Essays on practical reason and moral psychology*. Oxford: Oxford University Press.

Korsgaard, K. (2009). *Self-constitution: Agency, identity and integrity*. Oxford: Oxford University Press.

Lenman, J. (2013). Expressivism and constructivism. In Lenman & Shemmer (Eds.), *Constructivism in practical philosophy* (pp. 213–225). Oxford: Oxford University Press.

[21] See Tiffany (2012) for a nice statement of what I take to be a powerful dilemma for constitutivists. Smith responds directly to this dilemma in Smith (2015).

Ridge, M. (2005). Why must we treat humanity with respect? *European Journal of Analytic Philosophy, 1,* 57–73.

Ridge, M. (2012). Kantian constructivism: Something old, something new. In Lenman & Shemmer (Eds.), *Constructivism in practical philosophy* (pp. 138–158). Oxford: Oxford University Press.

Ridge, M. (2014). *Impassioned belief*. Oxford: Oxford University Press.

Ridge, M. (2015). I might be fundamentally mistaken. *Journal of Ethics and Social Philosophy, 9,* 1.

Scanlon, T. M. (2000). *What we owe to each other*. Cambridge, MA: Harvard University Press.

Schroeder, M. (2007). *Slaves of the passions*. New York: Oxford University Press.

Silverstein, M. (2012). Inescapability and normativity. *Journal of Ethics and Social Philosophy* (JesP). http://www.jesp.org/PDF/inescapability_and_normativity_final.pdf.

Silverstein, M. (2015). The shmagency question. *Philosophical Studies, 172,* 1127–1142.

Smith, M. (1994). *The moral problem*. New York: Wiley.

Smith, M. (2015). The magic of constitutivism. *American Philosophical Quarterly, 52,* 187–200.

Street, S. (2008). Constructivism about reasons. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 3, pp. 207–245). Oxford: Oxford University Press.

Tiffany, E. (2012). Why be an agent? *Australasian Journal of Philosophy., 90,* 223–233.

Velleman, D. (2009). *How we get along*. Cambridge: Cambridge University Press.

Wallace, J. (2013). Constructivism about normativity: Some pitfalls. In Lenman & Shemmer (Eds.), *Constructivism in practical philosophy* (pp. 18–40). Oxford: Oxford University Press.

Wedgwood, R. (2007). *The nature of normativity*. Oxford: Oxford University Press.

Williams, B. (1979). *Internal and external reasons. Reprinted in Moral Luck (1981)* (pp. 101–113). Cambridge: Cambridge University Press.