

# Kantian Constructivism and the Authority of Others

Aleksy Tarasenko-Struc

*Penultimate draft: please cite published version where available.*

**Abstract:** I argue that Christine Korsgaard's Kantian constructivism cannot accommodate our obligations to others. Because she holds that all of our obligations are grounded in our obligating ourselves, she is committed to the view that our obligations to others are grounded in corresponding obligations to ourselves. Yet this conclusion is objectionable on substantive moral grounds. The problem is that she embraces an egocentric conception of authority, on which we originally have the authority to obligate ourselves while others only have the authority to obligate us because we grant it to them. The solution is to adopt a more thoroughly social conception of authority and autonomy.

## 1. The Significance of Morality's Directedness

Many of morality's most significant requirements specify which forms of treatment we owe to other persons, not just how we ought to act where others are concerned. Intuitively, for example, we normally owe it to others not to kill them or cause them suffering; not to coerce, manipulate, or deceive them; not to interfere with their discretionary projects; to keep the promises that we have made to them; and to provide them with certain minimal forms of positive aid. These aren't just obligations *with respect to* or *involving* other persons, but obligations *to* them specifically.<sup>i</sup>

By all appearances, an obligation to another person takes a distinctive form, carrying a singular normative force. It is an obligation for two: a normative nexus joining a pair of distinct parties to one another. What makes such an obligation unique is that when we violate it, we also *wrong* the party to whom we are obligated; we don't just act wrongly with respect to some standard, value, or reason.<sup>ii</sup>

I will assume, I hope not too controversially, that there are obligations to other persons, that these obligations are largely what we take them to be, and that some of them are moral

obligations. I will also assume that the directed character of our moral obligations to others is essential to them. Taken together, however, these assumptions generate a criterion of adequacy on a moral theory. According to the *directedness constraint*, a moral theory is adequate only if it can be shown that the theory in question preserves the directedness of our obligations to others, so that these obligations are owed *to the right party, in the right way*—not, say, owed only to ourselves or to no one at all. Thus, if certain moral theories could be shown to violate the directedness constraint, this failure would give us reason to reject these theories for their inability to accommodate our obligations to others.

There is certainly a strong case to be made that versions of egoism and utilitarianism run afoul of the directedness constraint; I return to this matter below.<sup>iii</sup> My primary aim in this essay, however, is to show that a paradigmatic form of Kantian constructivism meets with a similar fate as well. My focus will be on Christine Korsgaard's version of the position.<sup>iv</sup> I will argue that Korsgaard's Kantian constructivism cannot meet the directedness constraint, and that its incapacity is due to its conception of the nature and grounds of our obligations generally.

Korsgaard's constructivism will serve as an ideal test case, for two reasons. First, not only is her view a classic of the genre, she has also drawn explicit attention to the directed character of our obligations and has made a concerted effort to demonstrate that the view preserves their directedness.<sup>v</sup> If Korsgaard's view does not meet the directedness constraint, then, and so cannot make good on this promissory note, we may have sufficient reason for rejecting this view, even by Korsgaard's lights. Second, if her view proves to lack the resources for accommodating our obligations to others, that raises the question of how general the problem is. If we demonstrate the inadequacy of a view whose aim is to vindicate the directedness of our obligations, that suggests that other moral theories may fail on this score, too. In other words, if

my argument against Korsgaard succeeds, it may be advisable to check our favorite theory to ensure that it does not share a similar defect.

So, after introducing key distinctions that demarcate the concept of an obligation to another person, I flesh out the directedness constraint and explain how a moral theory may violate it (§2). Then, I raise a basic challenge for Korsgaard's account of the publicity of practical reasons in light of her view of their ultimate ground. Since she holds that our obligations are grounded in the activity of obligating ourselves, it is unclear how we could have obligations to others—or, at least, how we could have obligations to others that are unmediated by prior obligations to ourselves (§3). After considering and rejecting a response to the challenge, I diagnose her failure as due to her embrace of an egocentric conception of authority, on which we originally have the authority to obligate only ourselves while others have the authority to obligate us derivatively, because we grant it to them (§4). Accommodating our obligations to others requires discarding this Cartesian picture of the relation between self and other, in favor a more thoroughly social conception of agency and autonomy, and I close by indicating, in very rough outline, what general features such an alternative should have (§5).

## **2. Preliminaries: Obligations to Others and the Directedness Constraint**

I begin with three clarificatory remarks about our obligations to others and how they are distinctive.

If I have an obligation to you, then it follows that in violating it, I wrong you specifically. My first point is that wronging is not equivalent to harming. Not all acts of harming wrong someone: think of just punishment. Perhaps not all acts of wronging harm someone, either, though I am unsure. At any rate, wronging another person has a normative significance not

shared by merely harming her. An act of wronging implicates or concerns the wronged party in a special way: it is directed at her in particular, as reflected by the fact that her wrongdoer is *accountable* to her for the way that he treats her. That is, if we have an obligation to someone, it follows that she enjoys a suite of normative powers: she has the authority to do such things as call on us to *justify* our actions to her, to *demand* certain conduct from us, to *blame* us if we do not comply, to call for *apology* or *reparations*, and so forth.<sup>vi</sup> Many of our moral obligations to others even correspond to *rights* that they have not to be mistreated in a variety of ways, and these rights collectively define the domain of interpersonal justice, or *right*.

Second, obligations to others contrast with obligations to ourselves, as well as with merely monadic obligations—if indeed there are any.<sup>vii</sup> I trust that the idea of an obligation to oneself is clear enough. A merely monadic obligation, on the other hand, is simply a requirement to act in a certain way, owed to no one; thus, no one would be wronged by its violation, no special normative powers would thereby be activated. One candidate here is our obligation not to destroy the natural environment. We seem to be under such an obligation, yet it may be that none of the Earth's inhabitants is wronged when I fell a redwood just to hear it crash to the ground, say; the planet itself is certainly not. If that is right, then what we have is a merely monadic obligation.

Third, an obligation that is not other-directed might still require acting for the benefit of others. Here I will say that we have an obligation *concerning* other persons, rather than an obligation to them. Admittedly, the distinction between an obligation to another party and an obligation concerning that party is not easy to characterize with precision, but I hope to convey its significance. On the surface, the difference is formal—a difference in normative structure, as Michael Thompson has pointed out.<sup>viii</sup> If Cain has an obligation to Abel not to kill him, that

obligation takes the form of a normative relation between those two individuals in particular, so that Cain wrongs Abel in violating that obligation. However, if Cain had a merely monadic obligation *concerning* Abel not to kill him, it would be a relation between one individual, Cain, and a general requirement to refrain from killing people—a requirement to which Abel would also stand, in a separate relation, so that he would be obligated not to kill Cain. Cain would not thereby count as wronging Abel in killing him, just as Abel would not if he killed Cain. Their obligations toward one another would then be on a par with their obligation not to destroy the natural environment, construed as above. If Cain cuts down a redwood to amuse himself, he violates this obligation and thereby acts wrongly. Yet there is a sense in which any given redwood constitutes ‘raw material for wrongdoing’, to borrow Thompson’s evocative phrase: the one that Cain fells is ‘the occasion, not the victim’ of his violation.<sup>ix</sup> If Cain’s obligation is merely monadic, it is an impersonal prohibition whose violation would not make a victim of anyone, Abel included—except in the anemic sense that he would be harmed by the offense.

Thus, obligations to other persons are importantly different from merely monadic obligations concerning others. They are also importantly different from obligations *to ourselves* concerning others. If Cain has an obligation only to himself not to kill Abel, Cain cannot wrong Abel but only himself. Again, Abel is just raw material for wrongdoing here, the occasion of the infraction but not its victim. In fact, if Cain’s obligation is only self-directed, then the true victim of Abel’s killing is *Cain*.

It matters to us whether we have obligations to others or merely obligations concerning them. Consider the obligation to keep our promises to others. Suppose, for example, that you have promised me that you will pick me up from the airport today, and that you have incurred an obligation to do so. If that obligation is merely monadic, there is a prior general requirement on

each person that she keep her promises to others, one instantiation of which is the specific requirement on you to fulfill your promise to pick me up from the airport. Yet, importantly, no one is wronged when you forget to come at the agreed-upon time, thereby leaving me stranded—not even me, despite my status as the promisee; once again, I figure merely as the occasion, not the victim, of your violation of the promissory obligation. Imagine, then, that I confront you later on, and you express remorse for failing to pick me up. If your obligation is merely monadic, then it seems that the content of your remorseful thought should be, in the first instance, *that you broke a promise to someone*, and only secondarily *that you broke a promise to me*.<sup>x</sup> And the same point holds if your promissory obligation is an obligation only to yourself, although your having such an obligation *does* entail that someone is wronged by your violation—namely, *you*; nevertheless, the immediate object of your remorse is still, that you broke a promise to someone. Yet, I submit, it *is* important to me that you feel consternation for breaking a promise, in the first instance, to me. I take myself to have a special status as a promisee in our promissory relationship, from which it follows that I do not enter your deliberations as an occasion for promise-keeping in the way that a particular redwood may enter my deliberations as an occasion for preserving the natural environment.

I trust that these remarks have clarified the target phenomenon enough to enable us to appreciate what is at stake in a moral theory's success or failure in meeting the directedness constraint. As I have said, to meet the constraint, a theory must demonstrably preserve the directedness of our obligations to others, so that these obligations turn out to be owed to the right party, in the right way. How might this constraint be violated? One way is by *entailing* that there are no obligations to particular other persons but only merely monadic obligations, obligations to

ourselves, obligations to some external authority (such as God), or the like. But few moral theories violate the directedness constraint in quite this way.<sup>xi</sup>

What is more common is that a view will run afoul of it by leaving an unbridgeable explanatory gap. Simple direct egoism presents a clear case. On this egoist view, the most basic requirement on us is the agent-relative and egocentric requirement that each of us maximize the satisfaction of her own preferences. But the requirement is a merely monadic obligation, it seems. Or, at best, it might be a self-directed obligation, owed to oneself. For even if Cain best satisfies his own preferences in not killing Abel (he avoids God's wrath, after all), how exactly could Cain have an obligation *to Abel* not to kill him, such that Cain would wrong Abel if he did in fact take his life? Indeed, if the kind of requirement given pride of place by the view is merely monadic or self-directed, how it is possible for us to have obligation to other persons?

The problem is not that the requirement to maximize the satisfaction of our own preferences fails to entail that there are any obligations to other persons. Whether there are any such obligations is presumably a substantive matter. The problem, rather, is that it is unclear how any obligation to another person *could* emerge from what seems utterly foreign to it, the (merely monadic or self-directed) basic requirement to satisfy our own preferences, given the sort of normative entity that each is. There seems to be no discernible link between our obligations to others and what, according to this form of egoism, are the fundamental constituents of the normative domain. So, there is an explanatory gap, and simple direct egoism does not seem to have the resources to bridge it.

The issue is not unique to an egoist position. Consider a familiar version of maximizing act-utilitarianism, on which the most basic requirement on us is the agent-neutral requirement to maximize the general welfare.<sup>xii</sup> Here, too, there's an apparent mismatch or incongruity between

this perfectly impersonal requirement, compliance with which seems to be owed to no one in particular, and bona fide obligations to others. How could the latter emerge from the former? For a utilitarian of this stripe, the normative domain contains only facts about the value (and disvalue) of states of affairs, along with the agent-neutral requirement to maximize the occurrence of valuable states of affairs. Given such agent-neutral, and thus merely monadic materials, it is unclear how there could be obligations to other persons to begin with. I want to emphasize that the issue is not that, according to such a view, it is not an *analytic* truth that we have obligations to others; the point is not a version of Moore's open question argument.<sup>xiii</sup> My point is not that the view fails to imply, as a logical or analytic matter, that there are obligations to others, but that its choice of basic normative currency makes their existence a total mystery.

Like simple direct egoism, maximizing act-utilitarianism leaves an explanatory gap that is likely unbridgeable, given features of its privileged requirement. It violates the directedness constraint, too, then. So, if our starting assumptions are correct—we have obligations to others, they are *essentially* obligations to others, and these are largely as they appear to be—both views turn out to be inadequate. That is not to say, however, that these views couldn't still demonstrate that the same kinds of actions that we appear to owe to others are, ultimately, ones that we ought to perform. Perhaps they could. Yet neither view seems capable of explaining how there could be obligations to others. Direct egoism and utilitarianism can only ever be extensionally adequate, not intensionally adequate, and in light of what we have assumed, their failure on this score gives us grounds for rejecting them.

These applications of the directedness constraint yield only modest results. For one, for all I have said, it may be that 'indirect' or 'two-tiered' versions of these views close the explanatory gap and meet the constraint; this possibility is not ruled out.<sup>xiv</sup> I point this out so as



to emphasize that the directedness constraint does not set an unreasonably high standard for a moral theory. The standard is only that a view not mislead us about basic features of our obligations to others, as, I have argued, simple direct egoism and maximizing act-utilitarianism do. Of course, the directedness constraint is easily met by what may be called *dikaiological fundamentalism*: the view that the fundamental constituents of the normative domain are obligations to other persons. But it is not only met by that view, and it remains to be seen whether more attractive views meet it. In any case, it certainly seems that certain features of a view's privileged normative entities readily create an unbridgeable explanatory gap. It is quite plausible, for instance, that the egocentricity of the egoist's principle and the agent-neutrality of the utilitarian principle, respectively, are what doom them. It seems fruitful, then, to assess other views' prospects for meeting the directedness constraint, considering the particulars of each case to draw out general lessons about how directedness may be undermined.

Determining whether familiar moral theories manage to meet the directedness constraint is a massive undertaking, to be sure, and I do not propose to complete it in a single paper. As part of this larger project, however, I turn now to an evaluation of Korsgaard's Kantian constructivism. The evaluation of other moral theories along this dimension will have to wait for a different occasion.

### **3. Korsgaard's Account of How Others Obligate Us: A Basic Challenge**

In *The Sources of Normativity*, Korsgaard sketches an account of how it is possible for us to obligate one another, one that is predicated on the idea that practical reasons are, in her terms, *public* rather than *private*.<sup>xv</sup> Reasons are 'public in their very essence', she claims, in that they are 'inherently shareable,' and their essential shareability reflects a particularly deep feature of

human beings: our social nature.<sup>xvi</sup> How should we understand her thesis that reasons are essentially shareable, and, in that sense, public?

As I interpret Korsgaard, her talk of the essential shareability of reasons for action refers to their capacity to figure in certain kinds of *interactions* among agents, as is evident in the following passage:

If I call out your name, I make you stop in your tracks... Now you cannot proceed as you did before. Oh, you can proceed, all right, but not just as you did before. For now if you walk on, you will be ignoring me and slighting me. It will probably be difficult for you, and you will have to muster a certain active resistance, a sense of rebellion. But why should you have to rebel against me? It is because I am a law to you. By calling out your name, I have obligated you. I have given you a reason to stop.

Of course that's overstated: you don't have to stop. You have reasons of your own, and you might decide, rightly or wrongly, that they outweigh the one I have given you. But that I have given you a reason is clear from the fact that, in ordinary circumstances, you will feel like giving me one back.<sup>xvii</sup>

This passage suggests that the shareability of practical reasons has two distinct but related aspects. First, the kind of interaction in which reason-sharing consists characteristically takes the form of one party giving another a reason to act in the way she is calling for—in the case above, a reason to stop—and it is standard for the recipient to decline to act in this way by articulating reasons of his own.<sup>xviii</sup> Thus, I suggest that for Korsgaard practical reasons are defined by their suitability for *co-deliberation*—that is, for the activity of reasoning together about what to do, in order to reach a shared decision. The constitutive aim of these interactions is to jointly arrive at a decision that both parties can basically accept—an aim that is particularly evident in contexts of coordination or collective deliberation, where the responses of our interlocutor are standardly taken as directly generating reasons for us as well. That the person with whom we are trying to schedule a meeting cannot make a certain time, say, is seen as giving us a nonstrategic reason to find a different time, too, one that is acceptable to us both.<sup>xix</sup> Second, when one person shares a

reason with another, it normally follows that the reason-recipient cannot simply ignore the reason in question without personally affronting or disrespecting the reason-giver.

Reasons for action are essentially public, then, in the sense that they are essentially shareable. And their shareability consists just in their capacity to be given by one party to another in the space of co-deliberation, so that, other things equal, ignoring them counts as disrespect for the reason-giver. A private reason, on the other hand, would be one that could not be exchanged between agents in this fashion; it would be a reason for only one agent. In Korsgaard's view, there are no private reasons. Accordingly, going forward, I will refer to public reasons, in this sense, more simply as 'reasons'.

In light of Korsgaard's conception of the publicity of reasons, then, there are grounds for taking her to be well positioned to accommodate the directedness of our obligations to others. For reasons, so construed, appear to bear the marks of directedness. They are necessarily such as to figure in co-deliberation with others, for one, and that is the exact context in which we would expect others to have the normative powers that, I have claimed, belong to the patient of a directed obligation. A partner in co-deliberation is, plausibly, the sort of party to whom we are accountable for abiding by our shared decision, and who is therefore normally entitled to call for us to justify ourselves to her, to blame us when we fail to comply, to forgive us when we have apologized or made amends, and so on.<sup>xx</sup>

Does Korsgaard's account meet the directedness constraint, then? Certainly, her account does not entail that we have no obligations to others, so it does not conspicuously violate the constraint. Does it leave an unbridgeable explanatory gap, however? To answer this question, we must examine her conception of the fundamental constituents of the normative domain, of their nature and ground. We will then be in a position to determine whether or not there is the

requisite sort of explanatory connection between our obligations to other persons and these allegedly fundamental constituents.

Korsgaard accepts a version of *metaethical constructivism*: the view that a practical requirement applies to a given rational agent, ultimately, in virtue in facts about that agent's practical perspective. On her view, a practical requirement applies to an agent in virtue of facts about his volitional activity—specifically, his engaging in the activity of self-binding, or, as I will also say, the activity of *obligating himself*. In an especially clear expression of her commitment to such a view, Korsgaard tells us:

the reflective structure of self-consciousness inevitably places us in a relation of authority over ourselves and... we are as a consequence also accountable to ourselves. By the reflective structure of self-consciousness, I mean the fact that we are conscious of the potential grounds of our beliefs and actions as such. When I am conscious that I am tempted to do something because of something else, I can ask myself whether I should do that, and this amounts to asking whether the consideration on which I propose to act provides a reason. To answer in the affirmative is to adopt a certain maxim of action as a law governing my conduct; to answer in the negative is to command myself to adopt a law of doing the opposite—refraining from the forbidden action and, if necessary, taking positive action to avoid the violation. Thus I act under my own authority as a lawgiver, and I am accountable to myself if I do not. So my reasons—and indeed practical reasons in general—are grounded in the authority the human mind necessarily has over itself.<sup>xxi</sup>

On this picture, an agent has a reason to  $\phi$  if and only if, and because, she has obligated herself to  $\phi$ .<sup>xxii</sup> I follow Korsgaard in using the term 'obligating oneself' in a broad sense, according to which an agent counts as obligating herself when she gives herself either a mere reason to  $\phi$  or an obligation to  $\phi$ . What it is for an agent to obligate herself to  $\phi$ , furthermore, just is for her to choose to  $\phi$  for the sake of some end E and to thereby adopt, as a universal law, the maxim of  $\phi$ -ing in order to achieve E.<sup>xxiii</sup> The ultimate ground of an agent's reasons is her authority to obligate herself, so construed. Because we exercise the relevant authority just by acting intentionally, this authority is essential to rational agency. Hence, if our obligations to others are

somehow constituted by (some subset of) our practical reasons, on Korsgaard's view, as I assume they are, these obligations are grounded in our obligating ourselves, and therefore in the kind of original authority that we as rational agents have to obligate ourselves.

Yet if the normative force of all of my obligations is in some way grounded in my obligating myself, and thus in my own authority to engage in self-binding, how is it that *you* can obligate me? And how can you obligate me in a way that yields obligations *to* you, not just obligations *concerning* you? To make the question concrete, suppose that while walking along Lake Michigan, I see you drowning. You call for my help. Because you are in peril and I can easily rescue you, I have an obligation to do so—indeed, an obligation to you to do so, so that I wrong you if I do not. So, you have obligated me. On Korsgaard's view, it would seem that I have an obligation to rescue you in virtue of my obligating myself to do so, where that consists in my choosing to pull you from the water in order to save your life and thus adopting, as a universal law, the maxim of helping others when they are in grave danger. My obligation to save your life seems to derive, in other words, from my authority to obligate *myself*. Does that explain why I have an obligation to you, though, and not just an obligation concerning you, so that I actually wrong you in failing to do so rather than merely acting wrongly with respect to you?

We may have reason to doubt that I could have a genuine obligation to you in the case above, if Korsgaard's picture of obligation is correct, for there *does* seem to be a problematic explanatory gap. The problem is that it is unclear how an obligation to you could emerge from my obligating myself, assuming that on Korsgaard's view all of my obligations are fully grounded in my so obligating myself, which rests, ultimately, on my having the corresponding background authority to so obligate myself. More generally, too, if our obligations derive wholly from our obligating ourselves, along with our background authority to obligate ourselves, none

of our obligations should plausibly be obligations to others but only obligations *to ourselves*, although some of these obligations will surely concern others.

This conclusion might be too quick, however. For my having an obligation to myself to rescue you need not exclude the prospect that aid is owed to you, too. After all, if I promise you that I will pick your friend up from the airport, I simultaneously incur two obligations: an obligation to you, and an obligation to your friend. In failing to do so, I wrong you *and* your friend, and that is reflected in the fact that both of you are entitled to blame me for violating that obligation.

Nor is this puzzling. At work in such a case, plausibly, is a familiar *transmission-of-obligation principle*. Roughly stated, the principle is as follows: if (1) one agent, A, obligates another agent, B, to  $\phi$  for the benefit of some third party, C, and (2) A's obligating B yields an obligation on B's part to  $\phi$ , then (3) B has an obligation to A to  $\phi$  and an obligation to C to  $\phi$ , so that she would wrong both parties through her noncompliance. Similarly, if the divine command theory is true, it may still be that my obligation to rescue you in the above case is an obligation to God and an obligation to you; both parties are wronged by my violation. And that is because my obligation, originally owed to God, is effectively transferred to you in light of His commanding all of humanity—hence, obligating me—to help others in situations of that kind.

It is therefore open to Korsgaard to insist that in obligating myself to save your life when you are drowning, I thereby generate two distinct obligations: an obligation to myself to save you and an obligation to you to do so. That is, she may appeal to the transmission-of-obligation principle adduced above, applying it to the limiting case in which a *single agent* occupies the place of both A and B. And she can affirm that some other-regarding obligations have a dual directedness: they are obligations to ourselves, since they arise through our obligating ourselves,

but they are also obligations to others. In that case, though, Korsgaard's Kantian constructivism contains no unbridgeable explanatory gap. For her conception of the fundamental constituents of the normative domain neither rules out the prospect of our having obligations to others nor makes the existence of these obligations a mystery. We can have obligations to others even if the ground of all obligations lies in our obligating ourselves.

Even if Korsgaard's view manages, in this fashion, to bridge the explanatory gap between our obligating ourselves and our having obligations to others, however, it does so only at significant cost. Suppose that the above account is sound: application of the transmission-of-obligation principle reveals that although our *obligating ourselves* is the ground of all of our obligations, these obligations need not always be obligations to ourselves alone but sometimes include *obligations to others* as well. Obligating oneself generates two distinct obligations, one self-directed and one other-directed. Thus, the absurd conclusion—that there are only obligations to the self—is avoided. But now we must ask: What is the relation between these two sets of obligations? It seems to be one of dependence: because my obligation to you is grounded in my obligating myself, it seems to follow not only that I have an obligation to myself as well but that the self-directed obligation is *prior* to the other-directed one. That is, from Korsgaard's view it seems to follow that whenever I have an obligation to another person, that obligation is *grounded* in an obligation to myself (with the same content), so that the former is always derivative of the latter.

This impression is confirmed by a casual examination of the cases considered so far. If I promise you that I will pick your friend up from the airport, it's true that I incur obligations both to you and to her, but it is also the case that my obligation to her is derivative of my obligation to you; they do not merely covary. Since it is you who have obligated me to do so, it is only in

virtue of my having an obligation to you that I also have an obligation to your friend; absent your obligating me, no obligation is generated. Similarly, again supposing the truth of the divine command theory, I have an obligation to you to save you from drowning; but since the obligation is created by God's obligating all of humanity to practice beneficence, I have that obligation to you only in virtue of having a prior obligation to God to do so. So, it stands to reason that if all of my obligations are fully grounded in my obligating myself, and in my background authority to do so, then my obligations to others are grounded in obligations to myself. But that is completely outrageous. Even if there is a sense in which I owe it to myself to rescue you, surely it cannot be the case that I owe positive aid to you *in virtue of* the fact that I owe it to myself—as reflected in the fact that if I had to rely on the thought that I had an obligation to myself in order to recognize that I have an obligation to you, I could well be accused of having one thought too many.<sup>xxiv</sup>

Would it be less outrageous for Korsgaard to affirm that my obligating myself (and thereby incurring an obligation to myself) is an *enabling condition* of my having an obligation to another person? Perhaps. On this sort of view, my obligating myself (thereby generating an obligation to myself) does not directly make it the case that I have an obligation to you; nevertheless, it does make it the case that the obligation's immediate grounds *make it the case* that I have that obligation to you.<sup>xxv</sup> This still strikes me as outrageous. Imagine a view—*simple promissory voluntarism*—on which an agent has an obligation to  $\phi$  if and only if, and because, she has promised herself that she would  $\phi$ . Suppose now that the promissory voluntarist makes the following claim: my promising myself to act for your benefit generates an obligation to you and an obligation to myself. The upshot would be that I have an obligation to you to save you from drowning in virtue of my obligation to myself to do so. Does simple promissory



voluntarism become that much more credible if it gives self-promising (and/or my promissory obligation to myself) the status of an enabling condition of my obligations to others?

I submit that it does not. This revised promissory voluntarism still implies that I have an obligation to you to save you *only if* I have an obligation to myself to save you. That implication is objectionable on substantive moral grounds. Just as it doesn't seem that I have an obligation to you to save you in virtue of my having the corresponding obligation to myself, it doesn't seem that I have that obligation to you only if I have an obligation to myself with the same content.

Notice that the problem with revised promissory voluntarism is not only that grounding our obligations to others in acts of self-promising makes these obligations contingent. For even if this enabling condition holds necessarily, even if we inescapably engage in self-promising (because, say, it is woven into rational agency), the obligations would still depend on the wrong sort of thing, just as they would if we had these obligations only when God commanded us to act accordingly. If we care in the right way about morality, we should not stake our moral concern on whether God has commanded us to act morally or, perhaps, on the matter of whether there is a God. Hence, in this case, it does not matter if it turns out that God necessarily issues these commands. If our obligating ourselves is an enabling condition of our obligations to others, on Korsgaard's Kantian constructivism, then it commits a similar error: it stakes our moral concern for others on whether we, in some substantial sense, obligate ourselves to act on the would-be claims of others (thereby accruing obligations to ourselves with the same content). But that is wrong. My obligations to others cannot be so dependent, whether directly or indirectly, on my obligating myself, much less on my having corresponding obligations to myself.

The directedness constraint requires that a moral theory preserve the directedness of our obligations to others, so that these obligations turn out to be owed to the right party, in the right

way. A theory violates this constraint if there is an unbridgeable explanatory gap between our obligations to others and whatever holds the status of fundamental constituent of the normative domain within the theory in question—that is, if it does not seem possible for the former to emerge from the latter. I have argued that Korsgaard’s view contains such an explanatory gap: it is unclear how we could have obligations to others if we suppose that all of our obligations derive from our obligating ourselves.<sup>xxvi</sup> That seems to suggest that there really are no obligations to others but only obligations to ourselves. Yet I have also conceded that the gap may not be unbridgeable, in light of a plausible transmission-of-obligation principle. The problem with relying on such a principle, however, is that it implies that we do have genuine obligations to others, but that these obligations are grounded in prior obligations to ourselves—a conclusion that, I have urged, is plainly objectionable on substantive moral grounds.

More generally, I have tried to show how the ground(s) of an obligation might determine whether it could be owed to someone in the right way. Certain grounds—such as purely egoistic considerations—will make our obligations to others objectionably dependent on conditions that are contingent and/or of dubious moral relevance. Korsgaard’s view, whether she conceives of self-binding as a (full or partial) ground or merely as an enabling condition, runs into a similar, problematic consequence: that our obligations to others are grounded in, or otherwise dependent on, our obligating ourselves or our corresponding obligations to ourselves. So, if Korsgaard’s brand of Kantian constructivism is true, our obligations to other persons may indeed be owed to the right parties, but these obligations will not be owed in the right way.<sup>xxvii</sup> Korsgaard’s view flouts the directedness constraint, then. And we have reason to reject it on this basis.

#### 4. Does Appealing to Joint Self-Binding Disarm the Challenge?

Thus far, I have assumed that Korsgaard holds that all of an agent's obligations derive from what may be called *individual self-binding*: his obligating himself singly, and thereby exercising his authority to do so. In more recent work, however, Korsgaard has clarified that an agent incurs at least some of his obligations when he and another party each reciprocally cede or transfer their authority to the other. I call this activity *joint self-binding*. Does appealing to the idea of joint self-binding enable Korsgaard to meet the directedness constraint? I will argue that it solves the letter of the problem but not its spirit.

As I use the term, two agents engage in joint self-binding when they transfer their authority to each other.<sup>xxviii</sup> According to Korsgaard, such a transfer may be effected by a promise or agreement.<sup>xxix</sup> Suppose that you and I have agreed to meet each other at a particular time. In that case, she says:

I give you authority over my will as to whether I will meet you at a certain time, provided that you give me authority over your will as to whether you will meet me at that same time, and in winning authority over the will that has authority over mine, I win myself back.<sup>xxx</sup>

In what sense do we transfer our authority to one another? Here the transaction in question seems to be a mutual transfer of what might be called *discretionary authority* with respect to some range of choices: an agent's right to unilaterally decide whether or not he will perform some action within that range.

Return to an earlier example: I promise you that I will pick you up from the airport today. So, before making that promise, I have discretionary authority as to whether or not I will take a walk later today, say, and my making the promise is my transferring that discretionary authority over to you. Once the promise is made, that is, you acquire the right to decide whether or not I

will go for a walk. As a right-holder, in other words, you enjoy a suite of corresponding normative powers: you're entitled to blame me if I do not comply, to demand compliance, and, crucially, to release me from the promise. And because you yourself have vowed to be at the airport at the agreed-upon time, you have, in turn, transferred to me your discretionary authority as to whether you will remain there then, so that I could legitimately hold you accountable if you simply left the gate before my arrival and hailed a taxi instead. Through this reciprocal ceding of authority, we commit ourselves to reaching shared decisions on these matters—decisions reached by joint deliberation, guided by the aim of basic mutual acceptability.

But recall Korsgaard's background commitment: according to her, the ultimate ground of an agent's obligations lies in his obligating himself, thereby exercising his authority to obligate himself. This suggests two points that she does not make explicitly but to which she appears to be committed. First, on her view, our obligating *one another* must consist in our obligating *ourselves* in tandem, and that must, in turn, consist in our reciprocally ceding discretionary authority to each other, as above.<sup>xxx1</sup> In such a case, in mutually transferring discretionary authority, I obligate myself to pick you up from the airport as you obligate yourself to meet me there, and in so doing, we obligate one another.

Second, for it to be possible for us to reciprocally cede our discretionary authority to each other through an act of agreement, it appears that another, prior transfer of authority must take place. Consider that even before I have promised to pick you up from the airport, you have some basic claim on me; for just by making your request, you commit us both to deliberating together, both about whether I will accept or decline that request, and if I do accept, about the particulars of the pick-up.<sup>xxxii</sup> How, though, can you have such a claim on me even before I cede my discretionary authority to you? A suitable explanation is that prior to our interaction *I have*

*granted you the authority to obligate me*; specifically, although I originally have the authority to obligate myself, I have transferred that authority to you—in effect, endowing you with the suite of normative powers that formerly belonged only to me. And, for your part, likewise, *you have granted me the authority to obligate you* in turn. That is, you have transferred to me an authority that originally belonged to you alone, the authority to obligate yourself. So, by engaging in joint self-binding, you and I have granted each other our own authority to obligate.

Why, then, might we think that Korsgaard's appeal to joint self-binding solves the problem that we encountered earlier?<sup>xxxiii</sup> Recall that problem is that her view seems unable to escape the conclusion that obligations to others derive from obligations to ourselves, that the latter are the ground (or the enabling condition) of the former. And that problem is generated by reliance on the transmission-of-obligations principle adduced above. The proposed solution would be to bypass this principle by embracing a *transmission-of-authority principle*. According to this alternative principle, if (1) one agent, A, transmits to another agent, B, his authority to obligate some third party, C, and (2) B's obligating C yields an obligation on C's part to  $\phi$ , then (3) C has an obligation to B to  $\phi$  but that obligation is not grounded in a more basic obligation to A to  $\phi$ .

Suppose, for example, that I'm organizing a conference, and that you—a graduate student in my department—agree to help me. In light of our agreement, I delegate a range of duties to you.<sup>xxxiv</sup> Specifically, I vest you with the authority to obligate your peers to perform certain conference-related tasks, e.g., collecting and reading submissions, send invitations to prospective speakers, and the like. Now, any other graduate student who fails to do what you have obligated her to do violates an obligation to you. Importantly, although this wayward student may thereby violate an obligation to me as well, it strikes me as bizarre to say that she owes compliance to

you *in virtue of* owing it to me. Since I have transmitted to you an authority that originally belonged to me, the authority to obligate your peers, it appears that any obligation to you that they may incur enjoys a kind of independent life. And we may also think that a similar conclusion holds of joint self-binding more generally. Thus, in granting you the authority to obligate me, thereby transferring to you my authority to obligate myself, any consequent obligations are owed directly to you; they are not grounded in obligations to myself.

Even with the backing of this transmission-of-authority principle, however, Korsgaard could appeal to joint self-binding and thereby, perhaps, win the battle but only at the price of losing the war. That is, if this principle is true, she may be able to escape the conclusion that our obligations to others derive from obligations to ourselves. Nevertheless, only the letter of the problem has been addressed. For on this elaboration of Korsgaard's view, it turns out that *you have no original authority to obligate me*: for you to have any authority with respect to me, I must grant you a share in my authority over myself.

In other words, her view is predicated on an *egocentric* conception of authority. It posits a problematic asymmetry between our authority over ourselves and the authority that others have with respect to us, so that whatever authority others have to obligate us must run through our authority, so to speak. But if we are justified in worrying, on substantive moral grounds, about the prospect of our *obligations to others* deriving from our obligations to ourselves, then the same grounds license us in worrying about the equally troubling prospect of the *authority of others* deriving from our own authority over ourselves. The general issue is that there is a dubious priority of self over other in the genesis of our obligations. Hence, the appeal to joint self-binding fails to solve the problem: the spirit of the problem is left intact.

Indeed, the problem is evident at various points in Korsgaard's discussion of self-binding.<sup>xxxv</sup> For instance, she holds that all interactions exhibit the normative structure that she ascribes to promising, treating promissory obligations as constituting the model of obligations more generally. But now notice that a promissory obligation to someone is only generated when the promise is actually made to her; that is, the obligation's existence depends on an act of voluntary ratification on our part. It is a distortion, however, to insist that all claim-making interactions work in that way, for the validity of certain kinds of claims on us does not seem to hang on our having assented to them, in any interesting sense.

This point is clearest in the rescue case above. If you are drowning and I can save you easily, you make a valid claim on me even when I don't first 'accept' your condition as reasonable and validate it—as if in your distress you were offering me, or anyone, a contract to sign and stamp. In other words, Korsgaard commits herself not only to an objectionably egocentric picture of how others obligate us but to an objectionably voluntaristic picture as well. In fact, her view of the nature and ground(s) of obligation is egocentric precisely in virtue of the kind of voluntarism essential to it—that is, in virtue of its holding that my obligations to you are always grounded in, or otherwise dependent on, an exercise of my own authority with respect to you.

That picture of our normative relations to others is unacceptable. It is, in fact, the same picture that is at the heart of simple promissory voluntarism: my promises to you are binding—that is, you acquire the standing of promisee—only if I promise myself that I will keep my promises to you, and you are in the corresponding position vis à vis me. So, there is a sense in which neither of us is in a position to make binding promises to the other *directly*. The normative force of any given promise to the other is always mediated by self-promising; without that initial act of self-promising on both our parts, none of our promises would be binding. Yet if our

promises to others are binding on these grounds, we don't stand in the right normative relation to others. Similarly, if others can obligate us only because we have transferred to them our own authority to obligate, then *a fortiori* we don't stand in the right normative relation to others, either. The authority of others does not derive from our own; we do not bestow that authority upon them.<sup>xxxvi</sup> Even if Korsgaard establishes that the relevant act of bestowal is one that we engage in necessarily, so that we inescapably share our authority with others, this is still to misrepresent our normative relations to others. For it is still to represent the authority of others as derivative in this way.

Korsgaard's Kantian constructivism fails to meet the directedness constraint, I conclude.<sup>xxxvii</sup> And the foregoing reflections suggest a plausible diagnosis of her view's inadequacy on this score, as due to her view of the grounds of obligation, which rests on the egocentric conception of authority.<sup>xxxviii</sup> A general lesson of my discussion, therefore, is that a view must reject this conception in order to be capable of meeting the directedness constraint. We must deny that the authority of others is derivative of our own. Others always already have the authority to obligate us and they very often obligate us directly. Beyond this negative recommendation, however, it is not easy to see our way to an alternative. Thus, I close with a suggestion about how we may arrive at a more suitable conception of authority.

## **5. Conclusion: Locating a Non-Egocentric Conception of Authority**

The egocentric conception of authority is the normative analogue of the Cartesian view of mind. Thus, I propose that understanding the weaknesses of latter will help us envision an alternative to the former.



The Cartesian view holds, roughly, that only our own minds are immediately accessible to us. While our standard way of knowing our own minds is non-inferential, we can only know facts about the mental states of others by making inferences from our experiences of their observable behavior. For the Cartesian, then, our own minds are directly available as objects of knowledge, yet the minds of others always lie at one remove. Others' mental states can never be present to us, properly speaking, although experience of their behavior can constitute evidence that they are in this or that mental state. The upshot is that we are profoundly isolated from one another. The mind of another person is a private realm about which we can form more or less well-substantiated conjectures, but these never approach the certainty with which we know our own minds. We are cut off from others, epistemically.

The egocentric conception of authority has a parallel structure and thus a similar upshot. Because any authority that you have to obligate me must derive from my authority to obligate myself, you can never obligate me directly; your claims have only an indirect normative grip on me, at best. In an important sense, we are cut off from others, practically, on this conception.<sup>xxxix</sup>

To locate an alternative to this picture, consider how we might respond to the Cartesian view of mind. Just as the proponent of the egocentric conception of authority must deny that others obligate us directly, the Cartesian denies that the minds of others are ever directly available to us as object of knowledge. Moving past the Cartesian view begins by acknowledging a sense in which the mental states of others *are* accessible to us directly—which is, however, different from our mode of accessing our own minds. The outward behavior of others *reveals* their mental condition; in particular, it doesn't constitute mere evidence on the basis of which we can infer that they are in some mental state or other. Hence, we retain a crucial aspect of the asymmetry between self-knowledge and our knowledge of others without admitting that we are

each locked away from one another in a private inner realm. Overcoming the Cartesian view requires us to conceive of the minds of others as inhabiting a public epistemic space.

Analogously, we should not regard each individual as a deontic monad, a self-contained locus of authority that may be rationally compelled, through some bridge principle, to share that authority with others. Others *already* share that authority with us originally, and no such principle is needed in order to bring them within our deontic orbit (and vice versa). We already occupy a public deontic space with others. Thus, just as a person's wince might be thought to directly reveal that she is in pain, the fact of her pain may likewise be thought to directly make a claim on us, where the validity of this claim in no way depends on our having validated it, or on our having granted her the authority to make claims on us more broadly. To make room for this alternative picture, we must reject certain other commitments of Korsgaard's view that create the impression that embrace of the egocentric conception of authority is mandatory.

More specifically, I propose that we discard Korsgaard's individualistic view of autonomy as well. To count as autonomous, on this sort of view, we must be moved to act by our appreciation of the normative force of a principle that we have imposed on ourselves—or, at least, of one that stands in a suitable relation to some duly self-imposed principle; otherwise, we count as heteronomous.<sup>x1</sup> Insofar as we are in the grip of this view, it will seem natural that we can only be bound by a principle validated by the free activity of our own will.

It is this sort of view that should be resisted. On the contrary, in affirming that other persons can obligate us directly, we must also insist that the claims of others—specifically, their needs, aims, and interests—can just as directly *move* us to act for their benefit. So, along with a more thoroughly social conception of authority, we need an equally social picture of autonomy,

so that it turns out that our being moved immediately by the claims of others does not constitute a threat to our freedom but rather an expression of it—and perhaps even a condition of it.<sup>xli</sup>

*Aleksy Tarasenko-Struc*

*Michigan State University,*

*East Lansing, MI*

*aleksyts@gmail.com*

## **Acknowledgments**

I presented versions (sometimes, distant ancestors) of this paper at a workshop at Harvard University, and at conferences at the University of Waterloo, Northwestern University, and the University of Pittsburgh. I benefited greatly from discussion with audiences at these conferences, from the responses of participants in an ethics reading group at Stanford University, and from the efforts of my commentators at Northwestern and Pittsburgh, respectively: Abigail Bruxvoort-Wilson and Jack Samuel. For particular helpful exchanges or comments, I would also like to thank Selim Berker, Susan Bencomo, James Bondarchuk, Thomas Breed, Matthew Boyle, Wesley Buckwalter, Byron Davies, Kyla Ebels-Duggan, Jeremy Fix, Micha Gläser, Chandler Hatch, Rusty Jones, Brendan De Kenessey, Douglas Kremm, Patricia Marechal, Daniel Muñoz, Oded Na'aman, Richard Moran, Thomas Scanlon, and an anonymous referee for the *European Journal of Philosophy*. I am most deeply grateful to Christine Korsgaard for her many insightful and generous comments on multiple drafts (and, well, for being a good sport).

## References

- Bagnoli, C. (2007). Respect and membership in the moral community. *Ethical Theory and Moral Practice*, 10, 113–128.
- Darwall, S. (2004). Respect and the second-person standpoint. *Proceedings and Addresses of the American Philosophical Association*, 78, 43–59.
- (2009). *The second-person standpoint: morality, respect, and accountability*. Cambridge, MA: Harvard University Press.
- (2013a). *Morality, authority, and law: essays in second-personal ethics I*. Oxford: Oxford University Press.
- (2013b). *Honor, history, and relationship: essays in second-personal ethics II*. Oxford: Oxford University Press.
- Ebels-Duggan, K. (2009). Moral community: escaping the ethical state of nature. *Philosophers' Imprint*, 9, 1–19.
- Enoch, D. (2011). Giving practical reasons. *Philosophers' Imprint*, 11, 1–22.
- Feinberg, J. (1970). The nature and value of rights. *The Journal of Value Inquiry*, 4, 243–260.
- Gaita, R. (1991). *Good and evil: an absolute conception*. New York, NY: Routledge.
- Gauthier, D. (1987). *Morals by agreement*. Oxford: Oxford University Press.
- Gilbert, M. (1993). Agreements, coercion, and obligation. *Ethics*, 103, 679–706.
- (1999). Obligation and joint commitment. *Utilitas*, 11, 143–163.
- (2006). *A theory of political obligation*. Oxford: Oxford University Press.
- Haase, M. (2014). Am I you?. *Philosophical Explorations*, 17, 358–371.
- Hare, R.M. (1982). *Moral thinking: its method, levels, and point*. Oxford: Oxford University Press.
- Kadlac, A. (2015). Does it matter whether we do wrong?. *Philosophical Studies*, 172, 2279–2298.
- Korsgaard, C. (1996a). The reasons we can share: an attack on the distinction between agent-relative and agent-neutral values. In *Creating the kingdom of ends* (pp. 275–310). Cambridge: Cambridge University Press.

- (1996b). *The sources of normativity*. Cambridge: Cambridge University Press.
- (2007). Autonomy and the second person within: a commentary on Stephen Darwall's *The second-person standpoint*. *Ethics*, 118, 8–23.
- (2009). *Self-constitution*. Oxford: Oxford University Press.
- Levinas, E. Freedom and command. (1987). In A. Lingis (Tr.), *Collected Philosophical Papers* (pp. 15–23). Dordrecht: Martinus Nijhoff Publishers.
- Moore, G.E. (1988). *Principia ethica*. Amherst, NY: Prometheus Books.
- Moran, R. (2018). *The exchange of words: speech, testimony, and intersubjectivity*. Oxford: Oxford University Press.
- O'Neill, O. (1996). *Towards justice and virtue: a constructive account of practical reasoning*. Cambridge: Cambridge University Press.
- Railton, P. (1984). Alienation, consequentialism, and the demands of morality. *Philosophy and Public Affairs*, 13, 134–171.
- Rawls, J. (1955). Two concepts of rules. *The Philosophical Review*, 64, 3–32.
- Reath, A. (2006). Self-legislation and duties to oneself. In *Agency and Autonomy in Kant's Moral Theory: Selected Essays* (pp. 231–249). Oxford: Oxford University Press.
- Schapiro, T. (2010). Desires as demands: how the second-person standpoint might be internal to reflective agency. *Philosophy and Phenomenological Research*, 81, 229–236.
- Smart, J.J.C and Williams, B. (1973). *Utilitarianism: for and against*. Cambridge: Cambridge University Press.
- Thompson, M. (2006). What is it to wrong someone? A puzzle about justice. In R.J. Wallace, P. Pettit, S. Scheffler, and M. Smith (Eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (pp. 333–384). Oxford: Oxford University Press.
- Williams, B. (1981). Persons, character and morality. In *Moral luck* (pp. 1–19). Cambridge: Cambridge University Press.

## Notes

---

- <sup>i</sup> These are also frequently known as ‘relational obligations’, ‘directed obligations’, or ‘bipolar obligations’.
- <sup>ii</sup> For simplicity’s sake, I focus on obligations to other *persons*, while leaving open that we have obligations to other animals as well.
- <sup>iii</sup> See §2, above.
- <sup>iv</sup> Korsgaard (1996a), (1996b, pp. 132–145), (2007), and (2009, pp. 177–206). Thus, I will not address the tricky interpretive question of whether Kant’s own moral theory contains the resources for meeting the directedness constraint.
- <sup>v</sup> Korsgaard (1996b p. 134). For a similar thought, see Korsgaard (2007, p. 11).
- <sup>vi</sup> In his work on the second person, Stephen Darwall has aptly highlighted the link between moral obligation, on the one hand, and relations of authority and accountability, on the other. See Darwall (2004), (2006), (2013a), and (2013b).
- <sup>vii</sup> The ‘merely’ is important, for every other-directed obligation has a monadic correlate. If I promise you that I will paint your fence, I incur an obligation to you to do so. But this obligation can be construed monadically, as the obligation to paint your fence (in light of my promise); this construal is silent on the question of whether or not the obligation is directed.
- <sup>viii</sup> Thompson (2006, pp. 335–338).
- <sup>ix</sup> Ibid. p. 352 and p. 340, respectively. See Gaita (1991 p. 148), along with chs. 4 and 9 in the same volume.
- <sup>x</sup> For a similar thought, see Kadlac (2014 p. 2283), and Gaita (1991 pp. 148–149).
- <sup>xi</sup> One example would be Joel Feinberg’s idea of a ‘sovereign right-monopoly’, in which all obligations are actually owed to ‘some “outside authority”,’ such as God. See Feinberg (1970, pp. 617–618).
- <sup>xii</sup> See, for example, the view defended by J.J.C. Smart in the first part of Smart and Williams (1973). On these direct utilitarian views, importantly, the requirement to maximize utility is agent-neutral and basic—not in the sense that it is ungrounded, but in the sense that it is a basic *requirement*. In light of these features, it is unclear why the requirement itself wouldn’t turn out to be merely monadic. Since the requirement to maximize utility (regardless of whose it is) is agent-neutral, no free-agent variable appears in the specification of any reason-generating fact countenanced by the view. But obligations to others are not only agent-relative but relative to two agents—dual-agent-relative, as it were. And because the requirement is basic, no other fact could explain why we owe it to others to maximize utility. So, there is an explanatory gap.
- <sup>xiii</sup> Moore (1988, pp. 15–17).
- <sup>xiv</sup> For an elaborate defense of two-tiered egoist theory, see Gauthier (1987). For prominent two-tiered utilitarian theories, see Rawls (1955), Hare (1981), and Railton (1984). Assessing the viability of these views is beyond the scope of this paper.
- <sup>xv</sup> Korsgaard (1996b, pp. 132–145), although it was preceded by Korsgaard (1996a).
- <sup>xvi</sup> Korsgaard (1996b, p. 135).
- <sup>xvii</sup> Ibid., p. 140. It’s worth noting that while interpersonal address is a paradigmatic mode of reason-sharing, on Korsgaard’s view, it is not the only one. We share reasons through non-verbal forms of communication (ringing a doorbell, telephoning); by pursuing ends with which others may be in a position to interfere (walking down the street); and even simply by being in certain conditions (like pain or peril) that we strongly want forestalled or ended. See *ibid.*, p. 141.

---

<sup>xviii</sup> The kind of reason-giving that Korsgaard appears to have in mind, particularly in the quoted passage, is closely related to what David Enoch calls ‘robust reason-giving’. Enoch claims that robust reason-giving is characteristic of speech acts such as demands, commands, and requests, and that it is capable of generating a new reason for action, not merely alerting someone to the existence of an already-existent reason. The case of calling out an acquaintance’s name seems to fit this characterization: even if I have no prior reason to stop and chat that is triggered by your presence, it still seems that I do have the relevant reason once you call my name, and that your speech act created this reason, in some sense. This is plausible for demands and commands as well. Contrast these cases of reason-giving with Korsgaard’s example of calling out for help, which does, in fact, appear to be more like informing someone of an already-existent reason to help. See, e.g., Enoch (2011).

<sup>xix</sup> Korsgaard (1996b, pp. 141–142). See also Korsgaard (2009, pp. 192–196).

<sup>xx</sup> Korsgaard (1996b, p. 141). Her discussion suggests that her notion of a public reason is broad, covering obligations to others and what might be called the mere reasons of others. The difference between them lies in their relative strength. If I have an obligation to someone to  $\varphi$ , it follows that I ought to  $\varphi$ , whereas if I have a mere reason to  $\varphi$ , it follows that I have only a *pro tanto* reason to  $\varphi$ ; it is not yet the case that I ought to  $\varphi$ . Some of Korsgaard’s examples are of mere reasons, in this sense; recall the reason given by the person who calls her name. Yet other cases are more appropriately classified as implicating full-fledged obligations to others. Thus, the reason grounded in someone’s peril is, properly speaking, an obligation to her, not a mere reason.

<sup>xxi</sup> Korsgaard (2007, pp. 10–11). See also Korsgaard (1996b, p. 104 and p. 165).

<sup>xxii</sup> Korsgaard (1996b, pp. 92–93, 104–105). For doubts about the idea that we literally obligate ourselves in the same sense in which others obligate us, see Haase (2014) and Moran (2018), ch. 7.

<sup>xxiii</sup> This terminology is clarified in Korsgaard (2009, pp. 11–12).

<sup>xxiv</sup> The idea is due to Bernard Williams, although he applies it in a different context. See Williams (1981, p. 18).

<sup>xxv</sup> Note that this proposal would not rely on the transmission-of-obligation principle. No transmission of obligations takes place, in fact: obligating myself to  $\varphi$  (thereby incurring an obligation to myself to  $\varphi$ ) only makes it the case that the immediate ground(s) of my obligation to you to  $\varphi$  (such as your needs and interests) make it the case that I have the latter obligation. In principle, the obligation to myself would simply be grounded in the very same thing—my obligating myself—that constitutes the enabling condition of my obligation to you. Here we do not seem to have a case of one obligation generating, or being transmitted through, another obligation.

<sup>xxvi</sup> Cf. Darwall (2006, ch. 9); he responds to Korsgaard’s argument on pp. 234–235fn.

<sup>xxvii</sup> Onora O’Neill’s Kantian constructivism invites a parallel challenge. According to O’Neill, all moral requirements, including principles of interpersonal justice, are authoritative in virtue of their capacity to coordinate action among a plurality of potentially interdependent rational agents with finite powers. (See, for example, O’Neill (1996).) From this conception, she claims, it follows that a central principle of justice prescribes the rejection of injury, so that we wrong others when our actions injure them directly or indirectly.

Once again, however, there is an explanatory gap. From the fact that a principle is not universalizable, that it is such that not all can coherently adopt it, it does not follow that anyone is wronged by an action performed on the basis of it. In fact, this conclusion *cannot* follow, for Kant’s universalizability test is supposed to generate both obligations owed to particular persons (perfect obligations) and undirected obligations to adopt general ends (imperfect obligations). We need some account of why a principle’s failure to be universalizable makes it so that we wrong someone in acting from it. And I do not see any such account in O’Neill’s work.

<sup>xxviii</sup> The discussion is scattered across Korsgaard (2009, pp. 186–202).

---

<sup>xxxix</sup> This is an instance of what Margaret Gilbert has called *joint commitment*. See Gilbert (1993), (1999), (2006), among other works.

<sup>xxx</sup> Korsgaard (2009, p. 193).

<sup>xxx</sup> *Ibid.*, pp. 190–193.

<sup>xxxii</sup> *Ibid.*

<sup>xxxiii</sup> Of course, the problem is not solved if we must *obligate ourselves* to grant one another the authority to obligate us. If I obligate myself to extend to you the authority to obligate me, I thereby create an obligation *to myself* to extend that authority to you, and once again, this self-directed obligation turns out to be more basic than the resulting obligation to you. Hence, we should suppose instead that, in instances of joint self-binding, my obligating myself *consists* partly in my granting you the authority to obligate me. Or perhaps, Korsgaard's view is that my obligating myself in these contexts assumes that I have *already* granted you the authority in question; indeed, Korsgaard suggests, in effect, that joint self-binding is a noumenal act of the will, so that in acting with others, we are always already engaging in joint self-binding with them. See her remarks on promises and agreements: *ibid.*, pp. 189–193.

<sup>xxxiv</sup> I am indebted to Kyla Ebels-Duggan for this example.

<sup>xxxv</sup> *Ibid.*, p. 189.

<sup>xxxvi</sup> Korsgaard's embrace of the egocentric conception of authority leads her to misconstrue the task of showing that our obligations to others are well founded. For she has suggested that confirmation that we have well-founded obligations to others lies in our experience of *our own* authority to obligate ourselves, which shows us that reasons are public; see Korsgaard (2007, p. 11). Tamar Schapiro, too, holds that the deliberative standpoint 'put[s] us into relation to an intrapersonal someone, a "you" within': we come to be aware of our capacity to act on directed claims through the experience of *desire*, which presents itself as a demand that our animal nature makes on our will. See Schapiro (2010, p. 234).

These views would still leave a gap, however. We originally have authority only over ourselves, but somehow the experience of recognizing that authority is supposed to make us aware that other persons can obligate us, and our authority over ourselves is also supposed to be the ground of the authority that others have with respect to us. Yet what is the connection, exactly? From the mere fact that we have authority over ourselves it does not follow that others have authority over us. Absent some further argument, then, the bare notion of our own authority over ourselves includes no basis for affirming others' authority with respect to us. There must be some normative bridge principle here. And the need for a bridge principle springs from the structure of the view itself so long as this view includes the egocentric conception of authority.

<sup>xxxvii</sup> I am perfectly willing to concede that there may be room, within Kantian constructivism, for a kind of no-priority view. Carla Bagnoli, for instance, advances a 'dialogical account' of respect, on which we reciprocally grant one another authority through 'an ideal dialogue'. This account, she says, 'takes seriously the claim that an adequate account should focus on the relation between the self and the community, but it does not hold the priority of either parties.' See Bagnoli (2007, especially p. 119). Because she follows Korsgaard's view very closely, I doubt that her approach would fare much better.

More promising is the view of Andrews Reath, who proposes that Kant's model of obligation be expanded: instead of just distinguishing who has the obligation (in his terms, the *subject*) from whom the obligation is owed to (the *source*), the model should also countenance a further role in the generation of the obligation—namely, the role of its *legislator*. According to Reath, 'no discrete individual is singled out as the legislator of a moral principle. While any moral agent must be able to identify with and to participate in the 'legislative process', it is not owned by any particular individual.' (Reath (2000), p. 241; see also p. 243.) Rather, the legislator of our obligations is none other than rational agents generally, engaged in shared deliberation to arrive at mutually acceptable principles. On this view, therefore, any given agent's obligations do not arise through the activity of her obligating herself but, in some sense, through all of humanity obligating itself.



---

If it could be shown to work, a no-priority view along these lines would be superior to a view like Korsgaard's. Yet such a view would face a variant of the challenge that I have raised against her view: it, too, seems to leave an explanatory gap. Since the view appears to hold that our obligations derive from all of humanity obligating itself, it must explain why our obligations to particular others do not turn out to be obligations to everyone. If it appeals to the transmission-of-obligation principle, it seems stuck with the consequence that our obligations to particular others are grounded in corresponding obligations to all of humanity—a consequence that would constitute a violation of the directedness constraint, too.

<sup>xxxviii</sup> The worry may apply to the metaethically ecumenical Kantian view offered by Kyla Ebels-Duggan as well. Ebels-Duggan argues that, given the extent of our interdependence, our autonomy can be maintained only if we acknowledge in others the authority to obligate us and vice versa. Even though her view is not constructivist, strictly speaking, she nevertheless tends to characterize the genesis of others' claims on us in voluntaristic terms. It is unclear whether she embraces the egocentric conception of authority or a no-priority view. See Ebels-Duggan (2009, pp. 8–18, but especially p. 14 and p. 17).

<sup>xxxix</sup> Actually, if the egocentric conception of authority is granted, it follows that we are cut off from others *epistemically* as well. For just as the Cartesian view of mind justifies me in questioning the reality of minds other than my own, the egocentric conception justifies me in questioning whether other persons have authority over me after all. Thus, in the latter case, too, the view itself creates the need for a separate argument to show that such a skeptical doubt can be defeated.

<sup>xl</sup> See, in particular, her discussion of Harriet in Korsgaard (2009, pp. 162–163).

<sup>xli</sup> For an articulation of a social conception of autonomy that has substantially informed my own, see Levinas (1987). In fact, the basis for a social conception of autonomy already exists in Korsgaard's work. For she holds that is our default stance to respond to the reasons of others as public, as reasons not only for the person whom they immediately concern but for ourselves, too: 'We do not seem to need a reason to take the reasons of others into account. We seem to need a reason not to.' See Korsgaard (1996b, p. 141). This form of interpersonal responsiveness runs deep with us, she thinks, making it so that '[we] can no more take the reasons of another to be mere pressure than [we] can take the language of another to be mere noise,' and that marks a profound sense in which human beings are social. If my argument succeeds, then Korsgaard would be right after all: we would have to be deeply social—indeed, even more so than she would have allowed.