EUJAP | VOL. 4 | No. 1 | 2008
ORIGINAL SCIENTIFIC PAPER
UDK: 17.02

COMMITMENTS OF A DIVIDED SELF: AUTHENTICITY, AUTONOMY AND CHANGE IN KORSGAARD'S ETHICS

LYDIA L. MOLAND
Colby College

ABSTRACT

Christine Korsgaard attempts to reinterpret Kantian ethics in a way that might alleviate Bernard Williams' famous worry that a man cannot save his drowning wife without determining impartially that he may do so. She does this by dividing a reflective self that chooses the commitments that make up an agent's practical identity from a self defined as a jumble of desires. An agent, she then argues, must act on the commitments chosen by the reflective self on pain of disintegration. Using Harry Frankfurt's emphasis on love as a final end, I argue that disintegration as motivation is not a more acceptable motivation than impartiality and so does not adequately address Williams' criticism. I also argue that the idea of a divided self either leads to an infinite regress or to an implausible description of how our commitments evolve and change. To make this last claim, I discuss a case from John Updike's novel *In the Beauty of the Lilies*.

Keywords: Christine Korsgaard, Harry Frankfurt, Bernard Williams, practical identity, Kantian ethics

Christine Korsgaard aspires, she tells us in *The Sources of Normativity*, to break down "Kant's overly harsh... division between natural impulses that do not belong to my proper self and rational impulses that do."[1] Meddling with Kant's distinction between which impulses belong to my proper self and which do not is no casual undertaking for a Kantian. What is at stake is nothing less than the link between the self and autonomy; for if the self is only a conglomeration of natural impulses, it is ruled instead of ruling, heteronomous rather than autonomous. It is, then, in the stricter sense not a *self* at all, not something that takes command of and directs its own life. If we are to be autonomous, we must guide our actions according to principles we rationally endorse rather than acting on unreflective desire or impulse. In order to be a self,

[1] Korsgaard 1996, p. 240. I would like to thank Bernard G. Prusak, Stephen White, David Wong, the participants in the workshop *Libertà e azione* at Università Roma Tre in spring 2006, and an anonymous reader for comments on earlier drafts of this paper.

in other words, I must act autonomously; acting autonomously means acting on principle.

But Kant interpreters have long complained that Kant goes too far, separating what we desire (our "natural impulses," as Korsgaard here puts it) from what we are ("my proper self") in a way that makes moral life unintelligible. To phrase the question in terms of Bernard Williams' famous example: if a husband must choose between saving two drowning people, one of whom is his wife and one of whom is a stranger, it would seem on a strict Kantian view that he must be impartial. Unless he can find a universal moral principle that can "legitimate his preference" (as Williams puts it) for saving his wife, we will be unable to describe his action as moral.[2] Williams maintains that the husband's search for such a principle involves "one thought too many," especially from the point of view of his wife, and so concludes that impartial Kantian morality is a "misrepresentation" of what moral evaluation should look like (Williams 1981, p. 19). Williams' example can be reframed to articulate a similar problem as regards the place of autonomy in Kant's ethics: if we define autonomy as resisting all natural impulses so as to act only on rational, universal principles, then the husband's saving his wife because of a natural impulse means he is not taking command of his own life and risks not being a *self*. But such a description of the place of persons that one loves in ethical decision-making surely skews our description of moral life. It portrays our most treasured relationships as obstacles to morality rather than central to it.

By conceding that Kant's division between natural impulses and rational impulses is "overly harsh," then, Korsgaard places herself in the tradition of Kant interpreters who hope to retain Kant's powerful link between autonomy and the self without claiming that actions motivated by natural impulses, even those impulses originating in generally morally laudable states such as love, by definition threaten the agent's autonomy.[3] How, then, to acknowledge the worth of some desires while maintaining autonomy in the Kantian sense? Korsgaard's answer is to describe a reflective self within each individual agent and then to differentiate it from the self made up of the jumble of desires we typically experience. This reflective self—through a process I recount below—chooses commitments that make up the agent's identity: commitments to friends, family, occupations, causes, and so forth. In so doing, it also chooses the desires that accompany those commitments. It gives itself these desires, in other words, and so remains autonomous. If the reflective self of the man in question has chosen his com-

[2] Williams 1981, p. 18. Williams borrows the example from Charles Fried in *An Anatomy of Virtues* (Cambridge: Harvard University Press, 1970). In stipulating that the man wants to save his wife because he loves her, I am already following another version of the example, namely Harry Frankfurt's in *The Reasons of Love*. I discuss this version of the example below.

[3] For other examples of this attempt, see for instance Allison 1990 especially pp. 184-191, where Allison defends Kant against a similar criticism as leveled by Allen Wood. See also 191-198, where he specifically addresses Williams, including a discussion of the drowning wife example. See also Baron 1984. Herman also gives an argument against this view in her 1993, ch. 9.

mitment to his wife, it will also have chosen his desire to save her if she is in danger. In this way, Korsgaard claims to be able to include our commitments and the desires they generate in the description of ethical life and, at the same time, to retain autonomy as the foundation of moral action.

But can postulating this divided self actually provide a description of ethical life that, by allowing and encouraging us to act on at least some desires, more accurately reflects our experience as ethical agents? Can it save Kantian-based ethics from the sorts of criticisms the drowning wife represents? I conclude below that it cannot, in part because of the way Korsgaard describes the division: Korsgaard's divided self, and especially her emphasis on the reflective self as the *real* self, ultimately prevents her from presenting a coherent theory of the self and its commitments. This becomes clear on two major counts. First: Korsgaard's description of the reflective self causes her to trace all obligation to the threat of the self's disintegration. This emphasis in turn causes her to misdescribe more pedestrian examples, ascribing to them a motivation no more palatable than the "overly harsh" description she seeks to mitigate. Second, the resulting theory cannot make sense of our experience of our commitments. It cannot explain how we actually adopt or relinquish a commitment and so what that commitment's real importance in our lives might be. In the first section below, I briefly summarize the argument that leads Korsgaard to postulate the divided self and describe its basic characteristics. In the second section, I detail the above criticisms, ultimately suggesting that Korsgaard's account leaves us still searching for the right way to describe ourselves and the place of our commitments in our agency.

1. The reflective self, the committed self

What is the nature of the divided self, and how are its components differentiated from each other? The answer comes in several steps. First: any human who experiences the sensation of choosing among various desires, Korsgaard claims, also experiences the sensation that who she is exists essentially separately or, as Korsgaard puts it, *above* those desires. This "something" over and above your desires is not just another, perhaps greater, desire, says Korsgaard. It is *you*, and that in a way that your other desires are not. Your determination of which desire to act on is then a product of who *you* are, not a product of one of the desires you contingently have: "This means that the principle or law by which you determine your actions is one that you regard as being expressive of *yourself*" (Korsgaard 1996, p. 100). This you is the reflective you, the part of you that makes you uniquely human (Korsgaard 1996, p. 92ff.). Having a part of yourself that independently chooses among desires allows you to be free in a way that you would not be were you to be determined by whatever desire external circumstances happened to awaken in you.

Korsgaard calls this self the reflective, or ephemeral, self.[4] (I will call the other self, which Korsgaard does not generally name, the committed self.) The reflective self is not free because of its interactions in the world: it is free because it is independent of the world. The capacity to distance yourself from whatever desires assail you is not developed through interactions with the world, but is presupposed in any human mind. A human being does not only desire, but evaluates her desires, reflecting on them and choosing which to act on. Through this reflection, we have evidence of her reflective self.

Being reflective presents us as humans with a problem. It "makes it both possible and necessary to decide which [impulses] we will act on: it forces us to act for reasons" (Korsgaard 1996, p. 113). The reflective you that acts from above your desires needs a coherent plan on which to act, to guide you through the "jumble" (as Korsgaard puts it) of competing desires. Korsgaard claims that these reasons are given to us by the various commitments we take up in the actual world. These commitments are in turn unified and ordered in what Korsgaard calls the agent's "practical identity." Practical identity is "a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking. . . . You are a human being, a woman or a man, an adherent of a certain religion, a member of an ethnic group, a member of a certain professional group, someone's lover or friend" (Korsgaard 1996, p. 101). I need a coherent set of reasons in order to act; that coherence is provided by my identity, my conception of myself in various roles. How I act is guided by this identity, by my desire to maintain an integrated sense of self. My practical identity gives me a plan on which to act and so fulfils my need, generated by my reflective nature, to act according to reasons.

Essential to Korsgaard's point, then, is that you are not only the jumble of these desires, or their sum. You are the reflective self behind them, and this you expresses itself through the coherent ordering of your various reasons and identities. Korsgaard enjoins us to consider "the astonishing but familiar 'I couldn't live with myself if I did that'." From this turn of phrase, Korsgaard concludes: "Clearly there are two selves here, me and the one I must live with and so must not fail" (Korsgaard 1996, p. 101).

The source of moral obligation, then, is in the division of the self into the reflective self and the committed self. If as a teacher I desire to fail a student because she annoys me, my reflective self will surmise that acting on this desire would be a breach of the reasons to act justly that my identity gives me. Part of being a teacher is the requirement that one grade one's students impartially. I will feel obligated to give my student a fair grade. If I desire to put my parents in a sub-par nursing home in order to save money for more glamorous vacations, my reflective self will recognize that such an action would break down my sense of myself as a daughter. Korsgaard writes that if living up to one's own standards is what makes someone "a person at all," then

[4] "Ephemeral" is what Korsgaard begins to call this aspect of the self in response to certain of Nagel's criticisms: see Korsgaard 1996, p. 229.

> To violate [those standards] is to lose your integrity, and so your identity, and to no longer be who you are. That is, it is to no longer be able to think of yourself under the description under which you value yourself and find your life to be worth living and your actions to be worth undertaking. It is to be for all practical purposes dead or worse than dead. (Korsgaard 1996, p. 102)

The wish to avoid this fate motivates us to choose those action that keep the integrity of our person intact:

> When an action cannot be performed without loss of some fundamental part of one's identity, and an agent could just as well be dead, then the obligation not to do it is unconditional and complete. Fulfilling the obligations of my practical identity is what I must do to preserve my identity and so my integrity. (Korsgaard 1996, p. 102)

This, then, is why I have the sensation of something else *above* my desires. My reflective self surveys my desires and determines which of them can be acted on without threatening my sense of integrity. The principle formed by my reflective self then obligates me to resist a desire that goes against my practical identity on pain of guilt and disintegration. A consistent pattern of such actions would result in a loss of self:

> So long as you remain committed to a role, and yet fail to meet the obligations it generates, you fail yourself as a human being, as well as failing in that role. And if you fail in all of your roles—if you live at random, without integrity or principle, then you will lose your grip on yourself as one who has any reason to live and to act at all. (Korsgaard 1996, p. 121)

Why does this theory count as an ethics of autonomy? My reflective self is giving my committed self laws when it insists that I not act on my desire to fail my student unfairly. In effect, then, I am giving myself laws: I am autonomous. As Korsgaard puts it: "Autonomy is commanding yourself to do what you think it would be a good idea to do, but that in turn depends on who you think you are" (Korsgaard 1996, p. 107). To put the point another way: one's practical identity mediates between the absolute need to act in a unified way and the question of what form that unification takes. What is demanded by my personhood is that I act coherently. The way I do that can be to be a teacher and so to treat my students well; to be a friend and so to exhibit qualities of trustworthiness, dependability, and so forth. In other words, "[a]n obligation always takes the form of a reaction against a threat of a loss of identity" (Korsgaard 1996, p. 102). I will have cause to challenge the extremity of this particular claim below; but for the moment, let us accept the suggestion that the way to avoid a loss of identity is to follow the laws and principles the reflective self produces.

The reflective self is, as we already saw, *me* in a sense that my other, committed self is not. It is above my desires, demanding that I hold a coherent identity together. It is not chosen in the way that many of our identity components are, nor is it something that is a result of our upbringing in a certain community. Being a reflective creature forces me to have reasons, to act as dictated by components of my practical identity: "But *this* reason for conforming to your particular practical identities is not a reason that *springs from* one of those particular practical identities. It is a reason that springs from your humanity itself, from your identity simply as *a human being*, a reflective animal who needs reasons to act and to live" (Korsgaard 1996, p. 121). In other words, my needing to act on some conception of practical identity is not, according to Korsgaard, a need I get from my practical identities. It is a need that is there already by virtue of my being human. This, then, is the answer to the question of why an agent *must* think of herself as having a practical identity. What the particular commitments of my practical identity are is, in an important sense, secondary. Particular commitments, Korsgaard says, can be "shed"; what cannot be shed is my need for those commitments (Korsgaard 1996, p. 120). The theory is, thus, very Kantian: one aspect of practical identity must be kept free of the contingencies of one's background or surroundings: it must be able to distance itself, look at the desires in question from a universal point of view. Only by orienting one's actions around this essential self can an agent be autonomous. Korsgaard writes:

> Most of the time, our reasons for action spring from our more contingent and local identities. But part of the normative force of those reasons springs from the value we place on ourselves as human beings who need such identities. In this way all value depends on the value of humanity; other forms of practical identity matter in part because humanity requires them. (Korsgaard 1996, p. 121)

This series of arguments has so far indicated that humans are reflective beings, and thus more than just their desires; that humans thereby need a practical identity; and that practical identity is based on our humanity. To maintain myself, to keep my unity intact, I must allow the part of me that is above my desires, reflecting on them from an impersonal point of view, to be in control. I must identify with this part of myself, realizing that it is this reflective capacity that allows me to maintain a coherent sense of identity and so of agency.

We can see how Korsgaard hopes through this model to blunt Kant's distinction between natural and rational impulses. If the husband has being a husband as part of his practical identity, then his reflective self has given him both this aspect of practical identity and the desires that naturally accompany it, for instance desires for his wife's well-being and survival. He need not in the moment reflect on what impartially is the best thing to do at the water's edge; he need not look for a rational impulse to justify his natural impulse. He can save his wife because she is his wife, thus avoiding any superfluous reasons and remaining autonomous.

Korsgaard makes clear, however, that her emphasis on the interplay between reason and desire is not meant to blunt the Kantian edge of her theory. There is still a fundamental insight of Kant's that she wants to preserve. Despite her desire, quoted above, to break down Kant's distinction between "impulses that do not belong to my proper self and rational impulses that do" (Korsgaard 1996, p. 240), Korsgaard writes: "Having said that, however, I want to return to, and emphasize, what is right about Kant's view that we should identify with our principles of choice rather than with our desires and impulses" (Korsgaard 1996, p. 240). What is right about this view is that "at the moment of action I must identify with my principle of choice if I am to regard myself as the *agent* of the action at all" (Korsgaard 1996, p. 240). In other words, Korsgaard wants to maintain the sharp distinction between the reflective self and the self that acts on desires. This dichotomy is clear when she describes an agent's reflective self as being the essential self, implying that the committed self is something less essential: "the principle or law by which you determine your actions is one that you regard as being expressive of *yourself*" (Korsgaard 1996, p. 100). She also describes the reflective self as having to take charge, "consciously [to] pick up the reins, and make *myself* the cause of what I do. And if I am to constitute *myself* as the cause of an action, then I must be able to distinguish between *my* causing the action and some desire or impulse that is 'in me' causing my body to act" (Korsgaard 1996, p. 227-228). I am, in the sense Korsgaard has explained, the author of my desires; but in order to preserve this authorship, I must always act as the reflective self who is behind the desires.

It is striking that, despite her careful attempt to suggest that desires and autonomy need not be incompatible, Korsgaard continues to maintain a sharp division within the self. It is a subtle difference, and necessary to state carefully. Korsgaard does not think that our particular commitments and the desires they produce harm our freedom. On the contrary, our free natures demand that we have practical identities in order to generate reasons and obligations on which we act. The desires generated by our practical identities are not damaging to our freedom, since at least they are desires arising from freely chosen roles. However, the way in which the committed self is free is fundamentally different from the way in which the reflective self is free. The ephemeral, reflective self is free already, without the roles it chooses for the committed self. In order to provide the necessary perspective that enables the committed self to act freely, it must be asocial, nonrelational, independent of the world. Since it can only act on reasons, it needs a self-conception to provide it with those reasons. But it is not these reasons that make it free, whereas the reasons the committed self gives itself *are* what make it free (insofar as it can be free at all, a claim I will challenge below). In that sense there is a radical disjoint between my reflective self and my committed self. There really is a split; there really are two selves.

Further: the freedom that is necessary for autonomous actions is still fundamentally separate from these commitments. Autonomy is anchored in the subject's relations to the world; but the controlling freedom, the reflective freedom that makes autonomous

action possible, is separated from the actual lives we live in the world. Practical identities then are attachments to and expressions of my reflective freedom, but have no real role in it. My being a teacher is part of my practical identity; I act autonomously through this role since my reflective self has given me the desires associated with being a teacher. But being a teacher does not really figure in the actual freedom that defines *myself*, as Korsgaard puts it. The freedom that defines *myself* is the abstract freedom of the reflective self.

Having seen how Korsgaard herself describes her adherence to and derivations from Kant's philosophy, it is worth pausing to consider other respects in which her theory differs from Kant's. First: several commentators have noted Korsgaard's move from a straightforward ethics of autonomy to an ethics of authenticity.[5] An ethics of autonomy need only show how I give myself the law; an ethics of authenticity is based on an analysis of who I *really* am. As Geuss puts it: "For Korsgaard I don't just *use* a criterion of formal law-likeness as a principle of endorsement of desires as reason for action, I '*identify*' with it."[6] I also, on Korsgaard's description, identify with components of my identity that articulate who I think I am in a more concrete sense—whether I am a pharmacist, sculptor, farmer, communist, German. To repeat: Korsgaard says that "Autonomy is commanding yourself to do what you think it would be a good idea to do, but that in turn depends on who you think you are" (Korsgaard 1996, p. 107). But it is not clear that Kant himself had any use for the question of which was the *real* self and how "who you think you are" might contribute to the coherence of the self.[7] Of course the agent's ability to have *her*self behind her actions is essential to Kantian autonomy: as Henry Allison puts it in a formulation parallel to the doctrine of transcendental apperception, "it must be possible for the 'I take' to accompany all my inclinations if they are to be 'mine' qua rational agent" (Allison 1990, p. 40). But this does not imply anything about being true to an authentic self who is comprised of specific commitments and must live up to those commitments on pain of disintegration.

Second, Korsgaard's claim that humans' reflective nature implies that "there are two selves here" is also not uncontroversially Kantian. Although several passages in Kant's

[5] See Skorupski 2000, p. 336, 351. This move is unKantian, according to Skorupski, since it derivates from Kant's appeal to the universality of reason and starts instead "from the modernist standpoint of authenticity—that is, from voluntarism about reasons and from the idea that one must construct one's identity and then do that which is required to maintain it" (p. 351). I think that Skorupski is correct that Korsgaard's existentialist interpretation of the self presents problems for her argument in general; my aim here, however, is not to recapitulate those arguments but to isolate the skewing effects this shift to authenticity has on our description of our commitments in particular.

[6] Geuss 1996, p. 191. Geuss continues with this line of criticism in the following pages. As Skorupski points out, Nagel also makes similar criticisms (Skorupski 2000, p. 351).

[7] Korsgaard denies Geuss' description of her existentialist depiction of the self as "unKantian." She quotes Kant as saying that "only as intelligence [is the person] his proper self" and as claiming that the agent does not attribute desires "to his proper self" (Korsgaard 1996, p. 237). As Skorupski points out, however, Kant is even in these comments not talking about a "radical or existential choice of normative identity" in the way Korsgaard is (see Skorupski 2000, p. 351).

works indeed suggest this ontological split, some Kant commentators argue that Kant's distinction between (in more Kantian terms) the empirical and intelligible character is not ontological but epistemic.[8] Instead of imagining our causally determined, empirical character as ontologically distinct from our free, spontaneous, intelligible character, we should ask what we must take to be true in order to make sense of our agency. As Allison puts it: "The basic idea is simply that it is a condition of the possibility of taking oneself as a rational agent . . . that one attribute such spontaneity to oneself" (Allison 1990, p. 45).[9] The idea of the dual characters is then a guide to how we should think of our actions rather than a description of distinct selves. [10]

I think these two points of Kant commentary are relevant for related reasons in Korsgaard's case. Because Korsgaard is articulating ethics in terms of authenticity, in terms of who I *really* am, she tends to use ontological language when discussing the self. Although she describes Kant's view as articulating a "*conception* of ourselves" that "does not prove the existence of a metaphysical self" but only asks us to act "as if there were something over and above all of your desires" (*SN*, 100), she also claims that "there really are two selves." She distinguishes between an "acting self" and an "empirical self" and between distinct "parts of me."[11] She describes the reflective self as being able to "shed," "give up" or "discard" particular commitments (Korsgaard 1996, pp. 120-121). Perhaps she does not mean these formulations to be definitive of her position, but they seem to suggest a view of a rational self separate and distinct from the committed self, able to reflect, evaluate and "shed" commitments independently of their importance in an actual life. That view of the self, I will argue, causes problems both for Korsgaard's description of the self and for her description of ethical life. Below, I articulate these criticisms, the first concerning the threat of disintegration and the second concerning the structure of the self.

2. Critique

A. Disintegration as motivation

Korsgaard, as we have seen, focuses key claims in her argument around an agent's fear that, should he not act according to a given part of his practical identity, he will lose his integrity and so his sense of himself. This focus on our fear of disintegration seems to me to yield results that are no better than the unduly harsh results Korsgaard tries

[8] Allison cites Terrence Irwin as an examples of an ontological reading. See Allison 1990, pp. 44-45.

[9] Allison also elucidates this distinction in his Introduction: see Allison 1990, pp. 1-6.

[10] In her 2002 Locke Lectures, Korsgaard phrases her description of the self in terms that more resemble Allison's reading: see http://www.people.fas.harvard.edu/~korsgaar/Korsgaard.LL1.pdf, p. 15. As I will argue, however, her descriptions of the self in *Sources of Normativity* suggest a more strict division.

[11] See Korsgaard 2006, pp. 104, 107, 231.

to avoid. To claim that if you violate your chosen standards, you lose "the description under which you value yourself and find your life to be worth living" and are "dead or worse than dead" is, if we consider our daily actions, certainly to overstate the case.[12] Perhaps in extreme situations we indeed face such extreme consequences. But this surely does not infer that every obligation-inspired action is an outgrowth of such acute fear. Aside from the fact that we often disregard our obligations without the fear of, let alone the serious threat of, disintegration (a fact Korsgaard acknowledges[13]), it is also true that much of the time we act according to our commitments because we want to, because it gives us joy, or simply because it is what we habitually do, what unproblematically belongs to the commitment itself. I doubt Korsgaard would deny this: she says, after all, that an obligation is unconditional only when an action involves violating "some fundamental part" of a person's identity. But she also writes, as we remember, that "[a]n obligation always takes the form of a reaction against a threat of a loss of identity" (Korsgaard 1996, p. 102). Korsgaard seems then committed to the idea that we at least subconsciously have the specter of disintegration before us at every moment, that every action is motivated ultimately by a fear of being dead or worse than dead.

Not only does this position seem descriptively false: Korsgaard's emphasis on disintegration means that her position fails also to improve on the overly harsh emphasis on impartiality Korsgaard earlier criticized in Kant. In an attempt to avoid describing actions as moral only when they are based on "rational impulses," her position instead depicts them as motivated by self-interest. The husband, because he is human, needs reasons on which to act. Being a husband gives him reasons to save his wife. True: the husband no longer saves his drowning wife out of principled duty; he saves her instead because he couldn't live with himself if he didn't—because he would disintegrate if he were to fail such a central commitment. Perhaps he truly couldn't live with himself if he didn't: but is that *why* he saves her? If it is, is that any improvement over saving her out of duty? If we accept that actions undertaken in the name of friendship, loyalty, family, dedication, and conviction are all at basis motivated by self-preservation, do we restore to those actions the natural, emotional, or instinctive quality that Kant seems to deny them? I think not: in which case Korsgaard's redescription of Kant's ethics does not achieve its goal.

The wife's complaint might also be reinforced by Korsgaard's description of what exactly makes the husband's commitment to her valuable. Our actual commitments are, to repeat, necessary since our humanity requires that we have reasons on which to act. This idea suggests, as Korsgaard says, that "other practical identities [than moral identity] depend for their normativity on the normativity of our human identity" (Kors-

[12] Both Cohen and Geuss make this point in the service of other criticisms: see Cohen 1996, p. 177; Geuss 1996 p. 195. Skorupski also for instance criticizes Korsgaard's claim that "it must be my project to preserve my self," concluding that since selfhood and reflective agency are not synonymous, one could be a reflective agent without being a self (Skorupski 2000, p. 345).

[13] Korsgaard 1996, p. 103.

gaard 1996, p. 125). Later she says that "moral identity is what makes it necessary to have other forms of practical identity, and they derive part of their importance, and so part of their normativity, from it. They are important in part because we need them" (Korsgaard 1996, p. 129). Again here, we can imagine the wife protesting at the idea that the importance and normativity of her husband's efforts on her behalf derive from his need to have a normative identity. Perhaps, she might say, this is not one reason too many but one identity too many.

In order to isolate the problem here, it is useful to compare Korsgaard's description of the threat of the self's disintegration to a similar claim made by Harry Frankfurt. Whereas Korsgaard describes humans as needing commitments since, being reflective, we must act for reasons, Frankfurt instead focuses on our fundamental need for final ends. Without final ends, Frankfurt writes, there "is nothing important for us to do," a state that can lead to boredom and threaten our "psychic survival" (Frankfurt 2004, p. 54). Frankfurt then claims that it is our *loves* that provide those final ends.[14] Love is valuable itself because it provides us with final ends, with things that matter to us in themselves. Our loves then gives us reasons to act in the interest of what we love. In doing so, loving allows us to sustain "the persistence and vitality of *the self*" (Frankfurt 2004, p. 55).

According to Frankfurt, the drowning wife's husband, like any other agent, needs final ends in order to maintain his self. Frankfurt (unlike Williams, who frames the example in terms of the relation of marriage) explicitly assumes that the husband loves his wife.[15] She is then one of his final ends. According to Frankfurt's account, she is instrumental in his maintaining of his self, not unlike the way in which on Korsgaard's view a commitment to one's wife can help constitute the self. Also under Frankfurt's account, if the husband fails to act according to his love, he risks disintegrating as a self.[16] Can we imagine the wife under Frankfurt's description protesting again that her husband has saved her only in self-interest? In what way, if any, is this an improvement on Korsgaard's description?

It seems to me that although Frankfurt's claim that we can lose our selves if we have no loves superficially resembles Korsgaard's claim that violating one's commitments

[14] Frankfurt also claims that what we love need not have value in any objective sense (see for instance Frankfurt 2004, p. 38), a claim Korsgaard explicitly challenges in her response to Frankfurt: see Korsgaard 2006. In the same essay, Korsgaard also contrasts her theory about the self needing practical reason to integrate the self with Frankfurt's claims about caring integrating the self: see p. 65. Frankfurt also discusses the value of love in his 1999, especially pp. 114ff.

[15] Frankfurt quibbles with Williams that it is not the legal status of marriage that should be in question here but that the "man in the example *loves* one (and not the other) of the two people who are drowning." The relation of marriage is not enough since it is possible that the man "detests and fears his wife" who "has recently engaged in several viciously determined attempts to murder him" (Frankfurt 2004, p. 37).

[16] See for instance his comments on Agamemnon in Frankfurt 1999, p. 139.

means a loss of self, Frankfurt is less open to the wife's complaint.[17] Instead of describing loves as things that a reflective self chooses in order to be a self, Frankfurt claims that the loves that are our final ends as not up to us (Frankfurt 2004, pp. 44, 49). The necessities these loves create "determine what [the individual] may be willing to accept as a reason for acting, what he cannot help considering to be a reason for acting, and what he cannot bring himself to count as a reason for acting. In these ways, they set the boundaries of his practical life" (Frankfurt 2004, p. 50). Thinking about final ends has then the following paradoxical air: "we may fairly say that final ends are instrumentally valuable just because they are terminally valuable" (Frankfurt 2004, p. 59). It is possible to be unhappy with what one is constrained to love; that belongs to our capacity as reflective beings. Conflicts among things we love can "put us at odds with ourselves" making it "impossible for us to plot a steady volitional course" (Frankfurt 2004, p. 50). We need not and indeed, because we are reflective, cannot simply accept our loves as they happen to us. But the starting place for agency on Frankfurt's view is these loves on which we then, in order to act autonomously, reflect.

By Frankfurt's telling, then, the wife is one of her husband's final ends. As such, he has *not* in fact simply *chosen* to love her, thus mitigating the fear that his commitment to her is at root an attempt to preserve his self. It is true that loving her fulfils an important part of his agency, but not because he has reflected rationally, concluding that loving her indeed fulfill the requirement that we act, to quote Korsgaard, "only on maxims that all rational beings could agree to act on together in a workable cooperative system" (Korsgaard 1996, p. 99). She can rest assured that he loves her as a final end.

What makes the wife's protests more plausible under Korsgaard's description than under Frankfurt's, I think again, is Korsgaard's language depicting a divided self, the reflective self surveying from a detached, rational point of view, choosing commitments, then assessing whether a given action is permissible in the "cooperative system" of the Kingdom of Ends (Korsgaard 1996, p. 99). Frankfurt's self is instead already formed by its loves, loves that are in turn "shaped by the universal exigencies of human life" (Frankfurt 2004, p. 47). In Frankfurt's view, the loves come first: whether one wills freely or not depends then on whether "what we want is what we want to want." If we indeed want what we want to want, "there is no respect in which we are being violated or defeated or coerced. Neither our desires nor the conduct to which they lead are imposed upon us without our consent or against our will." "Then," Frankfurt concludes, "we have on that occasion all the freedom for which finite creatures can reasonably hope" (Frankfurt 2006, p. 15).

[17] Frankfurt specifically imagines the wife's protests: "the lover seems inevitably to profit from, and hence to make use of, his beloved. Is it not clear, then, that love must inevitably be self-serving? How is it possible to avoid concluding that it can never be altogether selfless or disinterested?" (Frankfurt 2004, p. 60). Frankfurt calls this idea "perverse" since "[t]he fact that loving her is so important to him is entirely consistent with his being unequivocally wholehearted and selfless in his devotion to her interests" (p. 60). Frankfurt's own explanation for why his account would better satisfy the drowning wife involves also his claim that the husband under his description can save her "[w]ithout thinking at all" (Frankfurt 2004, p. 36).

B. The structure of the self

The division in the self suggested in Korsgaard's language also, I think, leads to a problematic image of the structure of the self. Korsgaard is right that I realize on some level that I am not just the sum of my commitments. The experience she calls obligation is indeed a familiar one: I am tempted to act in a certain way, but find that I am restrained by a sense of responsibility that extends beyond that particular desire. But does this imply the existential claim that who I *really* am, the myself that is in charge, must be a self "above and beyond" any given desire that then expresses itself through various contingent commitments?

I think not. Even if it were the case that my ability to act against my desires had a unique, essence-defining status for humans in general, it would not be clear how the resulting reflective self would be *my*self in particular. It has no distinguishing features that correspond to *my*self; it is simply free. Why then should *it* be the self that I "must not fail"? Korsgaard would certainly say that I must honor this self because it is the organizing principle of my life. But why should the organizing principle of my life be more me than my actual life?

There is, additionally, a problem if we try to decipher the way in which each self is free. Based on Korsgaard's description of the reflective self as acting on law and principle, it seems that the reflective self is free without interaction with the world. The committed self would then have the limited freedom of ordering its desires according to the reflective self's principles. But in order to order these desires, the committed self would need to be reflective as well. In that case, the committed self itself must be divided into a reflective self and a committed self, and an infinite regress looms. For how would the original committed self's committed self know to aspire to act on the ideals that the committed self's reflective self sets for it? If the committed self has any aspirations—which it must, in order to have the sense that it has something to live up to—it must also be reflective. But then how is it different from the reflective self?

Perhaps we should not imagine selves within selves, then, but rather another, greater me, a me that is *behind* both the reflective me and the desiring me, choosing (ideally) the reflective me. But by what criteria would *it* choose unless it was also reflective? Perhaps it is indeed reflective, and what we have been calling the reflective self is not in charge of choosing actions but is in charge only of constructing principles. Perhaps, in other words, there are three selves: the principle-making self, the committed, desiring self, and the reflective self that chooses between them. But which self would be living up to which self, in this case? Would the reflective self try to live up to the principle-making self? Which of the two would the committed self live up to? Again, how will the committed self aspire to live up to either if it is not reflective? The problems, it seems, proliferate as the selves do. Even if I could identify the reflective self as *my*self in some way, explaining how that reflective self interacts with the committed self continues to pose serious difficulties.

It seems to me, in other words, that Korsgaard cannot claim that the committed self is reflective, on pain of infinite regress. That leaves the committed self, as described, as the jumble of desires and the reflective self, the self acting on principle and law and *not* on the contingencies of the actual world, adjudicating all decisions about the desires the committed self is subject to. In other words, if we accept the distinct selves model suggested by Korsgaard's language isolating the "real" self, problems familiar from the Kantian case emerge.[18] If the empirical and intellectual characters are ontologically distinct, how do we explain the interaction between them? How can a self in time and a self out of time affect each other? If that is problematic, explaining why the *real* self is the intelligible or reflective self is all the more so. For what connection can there then be between my real self and my self in the world?

One of the images that most suggests a deeply split self is Korsgaard's repeated claim that we must be able to *shed* our commitments.[19] I have, under this description, multiple components of practical identity, but I must be able to shed them if reflection shows them to go against any of the reflective self's principles. I must be able to "discard" my identity as a daughter or a pharmacist or an American and the obligations that accompany these roles. I want to focus briefly on such cases as I think they isolate a particular problem with the ontologically divided self suggested by Korsgaard's theory. I will list three categories of change and indicate the difficulties each causes for this image of the self.

First: usually if I decide to break free of a commitment, the decision is the result of considerations as contingent as the commitment itself. I may switch political parties because I find the policies of my current party unethical; my concluding that they are unethical, however, will have to do with the news I hear, the way in which the policies in question affect me and others I care about, the current events at the time. In what sense can this decision then be made by a reflective self, given that this reflective self cannot be affected by contingencies? If I decide no longer to be a doctor, this will again have to do with my specific dislike of managed health care, with my desire for more humane working hours, with aspirations to take on another career. Again, these reasons are all contingent in the sense that *had* managed health care not existed or *had* family obligations not necessitated more time at home, I would not have rescinded my role as a doctor. Even if we focus on Korsgaard's insistence that we act on *principle*, surely most of our principles are chosen based on contingent experience as well. If my principle as concerns politics is to support political parties that combine economic growth with care for the least privileged, surely my experience of economics and social reality have influenced my support of that principle. If my chosen principle as a doctor is to give pa-

[18] Worries about how an intelligible self could interact with the empirical self are again at the heart of Allison's analysis and his preference for the epistemic reading of the division. See especially the section entitled "Timeless agency and the causality of reason" (Allison 1990, pp. 47ff.).

[19] See Korsgaard 1996, pp. 102, 120.

tients fair treatment, surely I have adopted that principle based on experience of what a doctor does, what patients need. But if the reflective self is free of such contingent considerations, it is again unclear how it could be involved in choosing or changing the principles that result.

Secondly, many if not most of our commitment changes happen gradually, without involving a particular moment of decision or renunciation. Close friendships die out, interest in causes or activities fades, life philosophies evolve. But it is difficult, in such cases, to pinpoint the role of the reflective self. Has it acted clandestinely, as something like a subconscious, weaning me of the commitment of which it, the reflective self, no longer approves? More likely, again, contingent factors intervene: my new significant other makes it difficult for me to spend the same amount of time with a friend and the friendship fades; my dedication to nuclear disarmament weakens because world hunger becomes a more pressing issue; my commitment to the belief that human nature is good transforms as I encounter evidence to the contrary. These causal factors are contingent: I could easily not have met my significant other, world hunger could never have become a problem or come to my attention, I could have encountered nicer people. As we look back over revisions in our practical identities over the years, probably more revisions are the result of this kind of contingent change than are the result of conscious decisions to drop or renounce a commitment. Given this fact, the reflective self as a separate entity again seems problematic: both because it cannot be influenced by contingencies and retain the kind of freedom that defines it and because change resulting from a non-conscious process seems incompatible with the nature of the reflective self. A reflective self cannot make gradual, non-conscious change since such changes are, almost by definition, influenced by contingent circumstances rather than rational calculation and evaluation. The fact that many changes indeed occur this way undermines, it seems to me, the picture Korsgaard presents.

We might argue in Korsgaard's defense that the reflective self is still active here, but only or only visibly in moments of reflection on these changes. I suddenly become aware that I have been neglecting my friendship because of my new romance. I remember that the self that I was trying to live up to before had maintaining friendships as an important part of its identity. Once reminded, I assess my more recent actions against that ideal. Two things can happen at this point. First, this moment of reflection might indeed provoke the sense that I have failed myself, and I might feel the threat of disintegration Korsgaard describes. In that case, I will either reinstigate my attention to my friend or disintegrate, and probably I will do the former. Second, I might decide that in fact my ideal concerning friendship is just not that important anymore. Not acting according to the ideal of friendship no longer threatens disintegration. The self that I must not fail now instead holds my romantic relationship up as something I must not fail. In this case, it seems again that the reflective self has changed in response to contingent information rather than evaluating actions against principles. Either that or the ideal it is holding the committed self to is very general and so can encompass both the

old, friendship-centered ideal and the new, relationship-centered ideal: an ideal such as treating all people with respect. If this were the ideal, we could imagine the individual then deciding that, although she does not spend as much time with her friend and the friend is hurt, she is not treating the friend with *dis*respect, exactly, and so pursuing her romantic interests instead is not a matter for personal disintegration. But if the role of the reflective self is this general, then Korsgaard's revision of Kant's categorical imperative does not in fact allow us to recognize the particularity of relationships instead of subsuming them all under a general moral law, as she claims. The reflective self cannot be *both* impervious to change *and* have ideals specific enough to address the particular relationships we form our ideals around.

Finally: even in cases in which an agent does suddenly shed a commitment almost in the manner of an animal molting out of a skin, I think Korsgaard's analysis is problematic. Consider the example of the Reverend Clarence Wilmot, first protagonist in John Updike's novel *In the Beauty of the Lilies*. Clarence, a middle-aged Presbyterian minister and father of three children, faithfully if wearily serves his diminishing congregation in Paterson, New Jersey. Clarence's religious belief functions as an aspect of practical identity in Korsgaard's sense: it gives him reasons for acting, lends coherence and purpose to his life, produces desires and goals. This is all the more true since, besides providing him with a religious affiliation, his belief underwrites his career and his social status.[20] The narrator informs us, however, that one day, without warning, "the Reverend Clarence Arthur Wilmot, down in the rectory of the Fourth Presbyterian Church at the corner of Straight Street and Broadway, felt the last particles of his faith leave him."[21] One moment Clarence's identity as a minister and a Christian is intact; the next, it is shattered.

The narrator is at pains to show us that no deliberation preceded this sudden reverse conversion: he reports that Clarence "was standing, at the moment of the ruinous pang, on the first floor of the rectory, wondering if in view of the heat he might remove his black serge jacket, since no visitor was scheduled to call until after dinnertime, when the Church Building Requirements Committee would arrive to torment him with its ambitions."[22] This initially arbitrary-seeming loss of faith is as absolute as it is sudden: "Clarence's mind was like a many-legged, wingless insect that had long and tediously been struggling to climb up the walls of a slick-walled porcelain basin; and now a sudden impatient wash of water swept it down into the drain." It comes to Clarence like a revelation: "*There is no God*."[23]

[20] In Frankfurt's terms, then, this would be a belief or ideal that Clarence loves.

[21] Updike 1996, p. 5.

[22] Updike 1996, p. 5.

[23] Updike 1996, pp. 5-6.

Clarence is, in a very real sense, not the person he used to be. He finds it quietly, despairingly extraordinary that when he writes, his handwriting is the same as it was when God, as he believed, existed; that the newspaper continues to deliver information from the outside world; that his wife continues to teach their Irish maid to cook like a Southerner. But Clarence immediately finds himself unable to comfort dying parishioners and literally loses his voice in the midst of sermons, as if his body were preventing him from perjuring himself. The refrain of God's non-existence continues to haunt him as he attempts to convince his wife, congregants and finally superiors that this is no momentary lapse of faith but an irrevocable reversal. When finally all accept that his faith is truly gone, his reasons for acting as he does, and indeed for keeping his job, are also gone. He works first in a department store and, after being laid off, as a door-to-door encyclopedia salesman. His family reluctantly, resentfully adjusts to his new directionless status. Not long afterwards, a prolonged case of tuberculosis first cripples and then kills him.

Certainly this portrait of a man rendered rudderless by losing a central component of practical identity supports several of Korsgaard's claims. First, it depicts someone worryingly close to being what Korsgaard calls a wanton, or someone who has lost the guiding thread that makes his life worth living and his projects worth undertaking.[24] It also supports Korsgaard's point that humans actually cannot operate without plan or goal: Clarence's dogged door-to-door salesmanship provides a practical identity, however insubstantial, that keeps him going. When his illness robs him even of this impoverished identity, he dies. Updike's depiction of Clarence is also pertinent to Korsgaard's theory, I think, because it comes close to the image of an agent shedding the commitments on which Korsgaard's description of the reflective self depends. Clarence recognizes essentially in an instant that the commitments that organized his life can no longer organize them. There is, to begin with, no contingently caused process or series of accidental realizations. It is in fact almost as if something outside of Clarence suddenly makes the decision and sheds the commitment. We could perhaps credit his reflective self, standing behind his particular commitments, with this realization.

But even in such a case as Clarence's, in which the simplicity of "shedding" or "discarding" or "giving up" makes sense, it seems untrue to the phenomenon to explain the change as the product of a reflective self who is, to repeat, in Korsgaard's theory separate from or above his current self: who adjudicates from a neutral space, needing *some* identity but not necessarily *this* one on which to act. The reason this does not seem right is, I think, also provided by Updike's character. Despite the suddenness and seeming unreflectiveness with which he becomes an atheist, Clarence himself explicitly seeks the reasons for this change in his background, personality, current situation, even the current political climate. He attempts to remember how he dealt with previous challenges to his faith; he remembers his father's wish that he become a minister and

[24] Again, Korsgaard's words: see Korsgaard 1996, p. 101.

blames his father for railroading him into a career he was, it now seems, unsuited for. He reconstructs his seminary days, searching for the point at which the seeds of doubt were sewn. The point is, simply, that he looks for, and finds, reasons for his lack of belief in the *particular person that he is*. His renunciation of his belief is not the decision of a being outside the belief, looking at Clarence's faith from a neutral point of view.[25] Who he is—his background, his personality, his feelings for his wife and family—constitutes the person who makes the decision to leave the ministry. Who he is after this decision fundamentally includes the abdication itself, along with his reaction to and reflection on the change. Clarence is, tragically almost, left to sort out a new practical identity: find new reasons for living, new ideas around which to organize his life, new ways of making the world cohere. That he never finds the same easy coherence that he had within his faith indeed reduces his sense of agency, leaving him self-conscious and unsure of himself. To repeat: that he nevertheless finds reasons, however pathetic, to go forward supports Korsgaard's description of the necessity of acting on *some* identity. What I doubt is that we can make sense of Korsgaard's claim that the way he chooses to go forward is mandated by an abstractly reflective self, the self Korsgaard's divide between reflective and committed selves suggests. If it is true, as I suggested, that the committed self cannot reflect and the reflective self is clearly distinct from the committed self and is responsible for all reflective decisions, Korsgaard is committed to claiming that Clarence's decision to leave the church is made by a self unbounded by particular commitments, clerical or otherwise. This, again, does not seem to me to be true to our experience. Perhaps Clarence is just wrong to look to these other factors to explain his decision: perhaps there really is a reflective self who has decided against his identity as a minister. But believing this would be to tamper radically with our understanding of how choices work and how our commitments fit together to form who we are.

It must surely be the case, as Korsgaard claims, that our autonomy is somehow bound up with our practical commitments and their correlative desires. On the one hand, autonomy without commitments leaves us the emptiness of the traditionally caricatured Kantian view, the demand that the subject abstract from all particular commitments in order to act morally. Commitments without autonomy, on the other hand, leave us fully determined by our surroundings and so not agents at all. And it surely must be true that our capacity for reflecting explains how we discern which commitments

[25] It is not my intention in this article to present a comprehensive defense of Frankfurt's theory, but Frankfurt's claim that we choose our loves and then are bereft when we find that the love in question cannot sustain reflection is, I think, a promising diagnosis of Clarence's disintegration. Frankfurt describes how we can find ourselves needing to revise our loves in several cases, one of which is that we find that "we do not understand the people and the ideals and the other things that we love well enough. Getting to know them better may reveal conflicts that previously were unnoticed. Our loving may turn out to have been misguided because its objects are not what we thought they were." Our inability to know ourselves completely adds to the possibilities for confusion: "It is not so easy for people to know what they really care about or what they truly love" (Frankfurt 2004, pp. 50-51, Frankfurt 2006 pp. 48-49). Again, it seems difficult to explain how the rational self as Korsgaard describes it could be mistaken or not know itself in this way, making Frankfurt's the more phenomenologically plausible of the two descriptions.

to take on and which to reject. It further seems right that when we feel obligated, we react to an aspect of ourselves that is beyond our desires. But to conclude from this, as Korsgaard does, that agency is made possible by a division between two selves is to prevent us from accurately describing ourselves and our commitments. The form this split self takes reduces every action to a bid for self-preservation, and so causes us again to misdescribe actions motivated for instance by love, friendship or enjoyment. Korsgaard's move towards authenticity—the claim the reflective me is the real me—leaves unclear what is *me* about me, since the reflective me can share none of my experiences, affinities or characteristics. Korsgaard's claim that there really are two selves combined with her image of our ability to shed our commitments prevents us from making sense of the myriad ways in which we adopt and relinquish commitments. Imagining such a reflective self downplays the historical narrative through which we understand our lives; it discounts the clues that narrative might provide as to why these commitments are so defining, so visceral, and can nevertheless sometimes be abandoned.

Korsgaard's goal is unquestionably a laudable one: to redescribe the place of commitments in an agent's autonomy so that the desires generated by those commitments do not interfere in the agent's freedom. But her efforts to combine the ethics of autonomy with the ethics of authenticity result in an image of the self that undermines the project's purpose. The split self may seem to secure the autonomy we require, but it does so at the cost of a coherent definition of the self and a phenomenologically accurate description of our relation to our commitments. It thus forces a distortion that goes to the heart of our ethical experience: namely, the experience of being motivated by the commitments we value most.

REFERENCES

Allison H. (1990), *Kant's Theory of Freedom,* Cambridge: Cambridge University Press

Baron, M. (1984), "The Alleged Moral Repugnance of Acting from Duty", *The Journal of Philosophy* 84, pp. 196-220

Cohen G.A. (1996), "Reason, Humanity, and the Moral Law", in O'Neil 1996, pp. 167-188

O'Neill O. ed. (1996), *The Sources of Normativity,* Cambridge: Cambridge University Press

Frankfurt H. (1999), *Necessity, Volition, and Love*, Cambridge: Cambridge University Press

Frankfurt H. (2004), *The Reasons of Love*, New Jersey: Princeton University Press

Frankfurt H. (2006), *Taking Ourselves Seriously and Getting it Right*, Stanford: Stanford University Press

Geuss R. (1996), "Morality and Identity", in O'Neill 1996, pp. 189-199

Herman B. (1993), *The Practice of Moral Autonomy,* Cambridge: Harvard University Press

Korsgaard C. (1996), *The Sources of Normativity*, Cambridge: Cambridge University Press

Korsgaard C. (2002), *Self-Constitution: Action, Identity, and Integrity*, Lecture One, "Metaphysical Foundations of Normativity," http://www.people.fas.harvard.edu/~korsgaar/Korsgaard.LL1.pdf

Korsgaard C. (2006) "Morality and the Logic of Caring", in Satz 2006, pp. 55-76

Satz D. ed. (2006), *Taking Ourselves Seriously and Getting it Right*, Stanford: Stanford University Press

Skorupski J. (2000), "Rescuing Moral Obligation", *European Journal of Philosophy*, *6*, pp. 335-355

Williams B. (1981), *Moral Luck*, Cambridge: Cambridge University Press

Received: June 15, 2006
Accepted: July 12, 2007

Colby College
4555 Mayflower Hill
Waterville, ME 04901
Lydia.Moland@colby.edu