Korsgaard's Expanded Regress Argument

Author: Samuel Kahn
Department of Philosophy, Indiana University-Purdue University, Indianapolis; Cavanaugh Hall 331a; 425 University Blvd; Indianapolis, IN 46202, USA
kahnsa@iupui.edu

Abstract: In this discussion note, I aim to reconstruct and assess Korsgaard's recent attempt to extend her regress argument. I begin, in section 1, with a brief recapitulation of the regress argument. Then, in section 2, I turn to the extension. I argue that the extension does not work because Korsgaard cannot rule out the possibility--a possibility for which there is both empirical evidence and argumentative pressure coming directly from the original regress--that we value animality in ourselves qua animality of rational beings.

Kant's ethics is notorious for the claim that, inasmuch as there are moral duties concerning nonrational animals, these are duties in regard to, and not to, these animals--duties that have their provenance in the psychological fact that behaving otherwise "weakens and gradually uproots a natural predisposition [in humans] that is very serviceable to morality."[1] The moral hierarchy Kant advocates is reflected in these well-known lines from the *Anthropology from a Pragmatic Point of View*:

> The fact that the human being can have the "I" in his representations raises him infinitely above all other living beings on earth. Because of this he is a *person*...i.e., through rank and dignity an entirely different being from *things*, such as irrational animals, with which one can do as one likes.[2]

Kant evidently thinks that there is nothing amiss in the idea that nonrational animals do not have moral standing and, therefore, that, considered in themselves, we may do with them as we will. But, others disagree. Thus, Pybus and Broadie assert that "Kantian rationalism must be rejected precisely because it is radically at odds with a sound ordinary moral view concerning our treatment of animals,"[3] and Bentham famously asserts that "[t]he question is not, Can they *reason*? [Kant], nor Can they *talk*? [Descartes] but, Can they *suffer*?"[4]

Recently, Korsgaard has sought to answer these objections. In particular, Korsgaard argues that, once we properly understand Kantian ethics, we shall see that it incorporates duties to (not merely in regard to) nonrational animals.[5] According to Korsgaard, her regress argument shows that rational nature is unconditionally valuable and ought to be respected in all of its instances--and, she claims, this regress argument can be extended to show that sentient nature is unconditionally valuable and ought to be respected in all of its instances.

In this discussion note, I aim to reconstruct and assess Korsgaard's project. I begin, in section 1, with a brief recapitulation of the regress argument. Then, in section 2, I turn to the extension. I argue that the extension does not work because Korsgaard cannot rule out the possibility--a possibility for which there is both empirical evidence and argumentative pressure coming directly from the original regress--that we value animality in ourselves *qua* animality of rational beings.

Section 1: The Regress Argument

Korsgaard's regress argument is put forward as a defense of the humanity formulation of Kant's Categorical Imperative, "act so that you use humanity, whether in yourself or any other person, always at the same time as an end, never merely as a means."[6] The regress argument may be reconstructed as follows:

(1) Rational agents necessarily represent their ends as objectively good.[7]
(2) Nothing is objectively good independently of rational beings.
(3) If (1) and (2), then rational agents necessarily represent their capacity for rational choice as unconditionally good.
(4) If rational agents necessarily represent their capacity for rational choice as unconditionally good, then rational agents always should treat humanity as an end in itself.
(5) Therefore, rational agents always should treat humanity as an end in itself.

I am going briefly to examine the premises of this argument before looking at how Korsgaard extends it into a derivation of duties to animals.

One way to understand premise 1 is by contrasting it with alternative theses in the problem space.[8] According to Korsgaard, X is objectively good if and only if X is unconditionally good, or X is conditionally good and the conditions of its goodness are satisfied.[9] For example, Korsgaard argues that happiness is objectively good if but only if the agent who is happy deserves to be so.[10] Now, some maintain that rational agents necessarily represent their ends as good, although perhaps not objectively good (and so agents might choose the lesser of two goods).[11] Others maintain that rational agents do not necessarily represent their ends as good at all.[12] These different alternatives offer (or rule out) alternative narratives of agents' intentional actions. Thus, according to premise (1), when agents say that they do not care about the fact that their ends are bad, the agents are either confused or lying (perhaps to themselves).[13] Korsgaard's argument for this premise is that, on the one hand, agents that are guided by reason pursue ends that are objectively good, and, on the other hand, whenever we act, we take ourselves to be guided by reason:

> [W]hen we act under the direction of reason, we pursue an end that is objectively good. But human beings, who act on their conception of laws, take themselves to act under the direction of reason. In the argument for the Formula of Humanity, as I understand it, Kant uses the premise that when we act we take ourselves to be acting reasonably and so we suppose that our end is, in his sense, objectively good.[14]

Premise (2) says that nothing is objectively good independently of rational beings. This is sometimes referred to as antirealism.[15] In Korsgaard's words, "our ends are not good in themselves, but only relative to our own interests."[16]

There is a subtle shift in this premise in Korsgaard's later work, when she begins to extend the regress argument. In particular, in her later work, Korsgaard asserts that nothing is objectively good independently of sentient beings. Thus, in *Fellow Creatures*, Korsgaard evinces her commitment to this modified version of premise (2) when she says that "[t]hings can be good or bad at all because they are good-for or bad-for creatures," where "creatures" is understood to include sentient, and not only rational, beings.[17] This will become more important in the next section of this paper.

Korsgaard's argument for this second premise is that, even if there is a non-natural property of objective goodness, we do not have access to it: "Without metaphysical insight into a realm of intrinsic values, all we have to go on is that some things are certainly good or bad *for* us."[18] That is, even if, metaphysically, there is some sort of real goodness out there in the world, the moral equivalent of radioactivity, we have no way of detecting it--there are no Geiger counters for goodness.[19]

Premises (2) and (3) often appear together in Korsgaard, as in the following:

> ...the things we desire have value because we want and need them, not the reverse. Our desire is a condition of their value. Our wanting them is not enough to make them good, however, for obviously many of the things we want are not good. Even if we want them we will not judge them good unless they are conducive to our happiness...Then we raise the further question...If we say

> it makes him happy, we ask why it is good that he should be happy…this condition will always be the presence of a good will.[20]

Korsgaard's idea is that, if all goodness is good for us rather than the reverse, then, when a rational agent represents her ends as objectively good, she presupposes that her capacity for rational choice is unconditionally good. This is where the regress appears: we have to perform a regress to determine the conditions of the objective goodness of our ends. Let me try to make this clearer with an example.

Suppose that I decide to eat a piece of chocolate. Then (from premise (1)), I represent that end (eating a piece of chocolate) as objectively good. But, plainly eating a piece of chocolate is not unconditionally good. So, I must take eating a piece of chocolate to be conditionally good and its conditions to be satisfied. A bit of reflection reveals that the conditions that make eating a piece of chocolate good begin with my having a corresponding inclination. But, Korsgaard argues that we do not take the satisfaction of inclinations to be unconditionally good: after all, satisfying some inclinations, like a gout sufferer's inclination for red meat, is not good because satisfying them will undermine, not contribute to, our happiness. So, the conditions that make eating this piece of chocolate good include the fact that to do so will satisfy an inclination that contributes to my happiness. But, according to Korsgaard, my happiness is not unconditionally good any more than satisfying an inclination is; my happiness is conditionally good, and its condition is my having a good will. That is, my happiness is not objectively good unless I deserve that happiness, and I deserve that happiness only if I am morally good.[21]

This is where the regress ends, according to Korsgaard. Thus, in taking my eating a piece of chocolate to be objectively good, I am presupposing that I have a good will and, further, that a good will is unconditionally good and the condition of the objective goodness of whatever I choose. Korsgaard evidently thinks that, in presupposing myself to have an unconditionally good will, I take my capacity for rational choice to be unconditionally good.

Korsgaard bases this last move (i.e., the move from the unconditional goodness of a good will to the unconditional goodness of the capacity for rational choice) on the idea that the capacity for rational choice and the good will are, in a sense, one and the same: the latter is merely the former perfected or successfully exercised.[22] The argument seems to be that, in taking ourselves to be successfully exercising our capacity for rational choice, we take ourselves successfully to be pursuing objectively good ends, and in pursuing objectively good ends, we take our capacity for rational choice to be unconditionally good, the condition of the objective goodness of our conditionally good ends.[23]

Premise (4) is based on the idea that we presuppose that the capacity for rational choice is unconditionally good *sans phrase*. I do not presuppose that my capacity for rational choice is unconditionally good *qua* capacity of someone who enjoys philosophy or someone who likes coffee or something. Thus, consistency requires that I treat this capacity with respect in all of its instances, and from this we may infer the conclusion that we ought to treat humanity as an end in itself.

Korsgaard's regress argument is controversial.[24] However, I am going to bypass this controversy in order to concentrate on Korsgaard's attempt to extend the regress into a derivation of duties to nonrational animals.

Section 2: The Regress Extension

Korsgaard, like Kant, thinks that animals are not rational.[25] She thinks that many animals are conscious and that their conscious states include pleasure, pain, fear, desire, emotions, beliefs (governed by causal association), and more. But, she thinks that animals are not aware of, and so are not able to assess, the grounds of their beliefs and actions. This, Korsgaard argues, entails that "animals are governed…by their instinct, desires, emotions, and personal attachments" in a way that we are not.[26] So, the regress argument does not, on its own, generate duties to animals. And this is where the regress extension comes in.

According to Korsgaard, there are two different ways to extend the regress argument. The first way, which she characterizes as weaker, focuses on the kinds of interests that we represent as objectively good. We are rational, but we are also animals. So, we have an animal nature and, accordingly, animal interests. Thus, Korsgaard

argues that "[p]art of what we confer value on when we respect ourselves is certain interests that we have, not as rational beings, but simply as sentient ones, such as the interest in avoiding suffering."[27] Consistency requires that we respect these interests wherever they are instantiated. So, it would be irrational to represent my own interest in avoiding suffering as objectively good while simultaneously representing the same interests of animals, like chickens or sows on a factory farm, as unimportant.

The second (stronger) way of extending the regress focuses on the kind of self on which value is conferred in the regress argument. Korsgaard asserts that, if we confer value on interests that we have *qua* sentient beings, then we are taking the sentient aspect of ourselves to be unconditionally good. Thus, the regress confers moral standing not only on beings with the capacity to set ends, but also on sentient beings. Consistency then requires that we treat all sentient beings, regardless of whether they are also rational beings, as unconditionally good:

> The stronger way to make the argument is just to say that because the original act of self-respect involves a decision to treat what is naturally good or bad for you as something good or bad objectively and normatively, the self on whom value is conferred is the self for whom things can be naturally good or bad. And the self for whom things can be naturally good or bad is your animal self: that is the morally significant thing we have in common with the other animals. It is on ourselves as possessors of a natural good, that is, on our animal selves, that we confer value. Since our legislation is universal, and confers value on animal nature, it follows that we will that all animals are to be treated as ends in themselves.[28]

Or, as she puts it in *Fellow Creatures*:

> As rational beings, we need to justify our actions, to think there are reasons for them. That requires us to suppose that some ends are worth pursuing, are absolutely good. Without metaphysical insight into a realm of intrinsic values, all we have to go on is that some things are certainly good-for or bad-for us. That then is the starting point from which we build up our system of values—we take those things to be good or bad absolutely—and in doing that we are taking ourselves to be ends in ourselves. But we are not the only beings for whom things can be good or bad; the other animals are no different from us in that respect. So we are committed to regarding all animals as ends in themselves.[29]

Korsgaard seems to favor the stronger version of the argument (with its focus on sentient beings rather than sentient interests).[30] But, she never decisively rejects one of these arguments in favor of the other. And because the objections I want to make apply to both arguments equally well, there does not seem to be any reason to choose between them.

Various respondents have suggested that we might value the interests/animality we share with other animals only insofar as they are the interests/animality of rational beings.[31] Because (as may be seen from the foregoing) constructivism is built on consistency requirements and explicitly disavows epistemic access to a metaphysical realm of Platonic values, this objection poses a serious problem. According to this objection, Korsgaard's extension of the regress generates a disjunction: treat sentient animals as ends in themselves, or claim that you value animality in yourself *qua* animality of a rational being.

Korsgaard offers two replies to this objection. One reply is that "anyone who made such a claim would be lying or engaged in self-deception."[32] Korsgaard challenges such a person to imagine that they are about to be deprived of their rational nature but, prior to being deprived in this way, they have the ability to determine whether they will be tortured afterward. Korsgaard thinks that nobody sincerely and reflectively would say that the torture should not matter.

Korsgaard's second reply is that the claim that I value my animality *qua* animality of a rational being "is not driven by the argument."[33] Korsgaard points out that there is no positive reason to think that I value my animality in this exclusive way. Moreover, she argues that many of the things we value are good for us not merely as

rational beings but also as animals. So, "it is more natural to think…that the presupposition behind rational choice is that animals…are ends in themselves."[34]

I think there are three problems with this.

First, Korsgaard's claim about sincerity is mistaken. To see why, consider Bernard Williams' famous thought experiments about personal identity.[35] You are told that you are going to have your brain wiped clean and, subsequently, to be tortured. Should you be afraid? Williams expects us to answer in the affirmative and argues that this confirms physiological theories of personal identity. The philosophical punch comes next: this thought experiment differs only in description, and not in substance, from a previous thought experiment that is supposed to elicit an intuition that confirms psychological theories of personal identity. But, in my experience, Williams' punch does not connect. Even on the best of days, only a small handful of my students answer in the way that Williams expects them to when asked whether they should be afraid (and that is only when I do not emphasize that the question is not whether they *would* be afraid). The reason this is relevant is that having your brain wiped clean is probably just about the same, for the purposes of this kind of thought experiment, as being deprived of your rational nature. So, I think there is every reason to believe that many people would say, sincerely and reflectively, that it should not matter whether they are tortured after they are deprived of their rationality.

To be clear, I do not think that erasing memories and erasing rationality are the same. In fact, when I run through Williams' thought experiments in class, I generally emphasize to students that, although they would lose their current memories, they would retain all of their reasoning and other cognitive capacities (including the capacity to form new memories of the pain that the torturer is about to inflict). Nonetheless, I think that, if people sincerely are indifferent to post-memory-erasure-torture, then people also can be sincerely indifferent to post-rationality-erasure-torture. Why so? Well, although I think I can imagine a nonrational being with memory (some fish might fit this description), it is hard for me to imagine a rational being entirely losing its rational capacities without losing access, or at least a certain kind of access, like the ability to form extended and meaningful narratives, to many, if not most, of their memories. So, it seems to me that the total loss of rationality is a more profound loss than the total loss of memory: the former seems to entail some sort of memory loss, whereas the converse of this seems more tenuous--and, even if I am ultimately mistaken about this, the *seeming* is sufficient to reinforce the argument I am making, that the fact that students sincerely are indifferent to the threat of post-memory-erasure-torture lends plausibility to the idea that people sincerely can be indifferent to the threat of post-rationality-erasure-torture. Remember, after all, that the question here is not whether people could be making a true assertion about this but, rather, whether people could be making a truth*ful* assertion about this, and, based on the foregoing, that seems eminently possible to me.

Now Korsgaard might concede this point but argue that (i) her argument nonetheless goes through for those who do think it should matter if they are tortured after they being deprived of their rationality, and (ii) there are other, perhaps moral, reasons to try to make oneself into a person who thinks such torture should matter. This leads me to the second thing I would like to say.

I think that there is the potential for confounding here. To see why, note that many people care about whether their corpses are defiled after they die.[36] I think it would be a mistake to infer that this is because such people value corpses as ends in themselves *sans phrase*. Indeed, in many parts of the Western world and for many years, the most serious crimes would be punished with death and, then, post-execution "enhancements." A person would be killed, and then their body might be torn apart by horses, the pieces displayed prominently on pikes in different places. This was in part because of the belief that, if someone's body was dismembered, they would not be able to participate in the resurrection. The reason this is relevant is that it suggests that many people care about whether corpses are defiled precisely because they care about the *particular* corpse (e.g., they care about whether *that person* will be able to participate in the resurrection). This poses a problem for Korsgaard because it lends credence to the idea that, insofar as we value our animal interests (the interests of our body while the latter is still alive), we do so *qua* the interests of a privileged group of animals and *not* qua the interests of animals *sans phrase*. Further evidence for this might be drawn from the widespread complaisance in the face of the fact that there are no provisions for animals in the afterlife in most religious traditions--or, to appeal to more contemporary culture, from the widespread complaisance in the face of practices on meat farms, fur farms, dairy

farms, and egg farms, and from the widespread complaisance in the face of practices in the entertainment industry, in the cosmetic industry, in the pharmaceutical industry, and in the research industry.

Now, some might object at this juncture. They might concede that, as a matter of fact, we seem to value our animality *qua* the animality of a rational being. But, they might argue that this is irrelevant: Korsgaard′s argument is about consistency, and so it is about how we *should* value our animality. So, (the objection concludes) even if we *do* value our animality *qua* animality of a rational being, Korsgaard′s argument shows that we *ought* to value our animality *sans phrase*, and that is the end of the story.

The problem with this objection, however, is that it has mistaken the conclusion of Korsgaard′s argument. As noted above, the conclusion of Korsgaard′s regress extension is actually a disjunction: we ought to value our animality *sans phrase* (and, therefore, we ought to treat animals as ends in themselves), or we ought to value our animality *qua* animality of a rational being. This is precisely why Korsgaard is so adamant about the claim that anyone who takes the latter route would be lying or engaged in self-deception–if someone sincerely takes this latter route (not engaged in lying or self-deception), then there is no consistency requirement for them to treat animals as ends in themselves. And this is also precisely why it is so problematic for Korsgaard that, as a matter of contingent empirical fact, there is an abundance of evidence that many people seem to value their animality *qua* animality of a rational being rather than *sans phrase*. This leads me to the third and final thing I want to say.

Korsgaard′s extension threatens to undercut the original regress on which it is built. In Korsgaard′s own words, her regress argument is supposed to show that "the unconditioned condition of the goodness of anything is…the power of rational choice."[37] It follows immediately from this conclusion that the power of rational choice is the condition of the objective goodness of choices regarding animal interests. So, these animal interests (and the corresponding selves that generate them) cannot have the same status as the rational nature that confers their status on them. This reveals a deeper problem: it looks like the regress argument leads us naturally to the conclusion that animal interests/selves have value insofar as they are the interests/selves of rational beings, exactly what Korsgaard is keen to deny. That is, when we remember the original regress on which the extension is built, we can see that the claim that we value our animality *qua* animality-of-rational beings is, in fact, precisely what is driven by the argument (*pace* Korsgaard).

I do not want to conclude from this that Kant′s or Kantian ethics is foiled by the problem of its treatment of nonrational nature. There are many other approaches (other than Korsgaard′s) in the literature on this topic, and some, like Denis, Ripstein, Tenenbaum, and Timmermann, even argue that the moral phenomena can be saved using a traditional Kantian framework.[38] But, I do want to conclude that the regress extension, although ingenious, is fatally flawed.

Bibliography

Allison, Henry. 1990. *Kant's Theory of Freedom*. Cambridge University Press.

Bentham, Jeremy. 1907. *An Introduction to the Principles of Morals and Legislation*. Clarendon Press.

Bergelson, Vera. 2009. "Consent to Harm." In *The Ethics of Consent* (ed. Franklin Miller and Alan Wertheimer), Oxford University Press: 163-192.

Birch, Jonathan. 2020. "The Place of Animals in Kantian Ethics." *Biology and Philosophy* Vol. 35 Issue 8: 1-11.

Chignell, Andrew. 2020a. "Book Review." *Mind*. December: 1-11.

Chignell, Andrew. 2020b. "Hope and Despair at the Kantian Chicken Factory." In *Kant and Animals* (ed. John Callanan and Lucy Allais), Oxford University Press: 213-238.

Denis, Lara. 2000. "Kant's Conception of Duties Regarding Animals." *History of Philosophy Quarterly* Vol. 17: 405-423.

Engstrom, Stephen. 1992. "The Concept of the Highest Good in Kant's Moral Theory." *Philosophy and Phenomenological Research* Vol. 52: 747-780.

Garthoff, Jon. 2020. "Against the Construction of Animal Ethical Standing." In *Kant and Animals* (ed. John Callanan and Lucy Allais), Oxford University Press: 191-212.

Gewirth, Alan. 1978. *Reason and Morality*. Chicago University Press.

Godfrey-Smith, Peter. 2021. "Philosophers and Other Animals." *Aeon*: 1-14.

Herman, Barbara. 1993. *The Practice of Moral Judgment*. Harvard University Press.

Hill, Jr., Thomas E. 2002. *Human Welfare and Moral Worth*. Clarendon Press.

Kerstein, Samuel J. 2006. "Deriving the Formula of Humanity (GMS II, 427-437)." In *Groundwork for the Metaphysics of Morals* (ed. Christoph Horn and Dieter Schönecker), Walter de Gruyter: 200-224.

Kahn, Samuel. 2013. "The Guise of the Objectively Good." *The Journal of Value Inquiry* Vol. 47: 87-99.

Kahn, Samuel. 2017. "Reassessing the Foundation of Korsgaard's Approach to Ethics." *Dialegesthai Rivista di Filosofia*. < **https://mondodomani.org/dialegesthai/articoli/samuel-kahn-01#fn:29**> Accessed 2021.06.09.

Kahn, Samuel. 2019. "The Problem of the Kantian Line." *International Philosophical Quarterly* Vol. 59, Issue 2: 193-217.

Kahn, Samuel. 2021. *Kant's Theory of Conscience*. Cambridge University Press.

Korsgaard, Christine. 1996a. *Creating the Kingdom of Ends*. Cambridge University Press.

Korsgaard, Christine. 1996b. *The Sources of Normativity*. Cambridge University Press.

Korsgaard, Christine. 2004. *Fellow Creatures*. Tanner Lectures on human Values 24: 77-110.

Korsgaard, Christine. 2008. "Just Like All the Other Animals of the Earth." *Harvard Divinity Bulletin* Vol. 36, No. 3: 1-8.

Korsgaard, Christine. 2009a. "Exploiting Animals." *AV Magazine* Fall: 14-15.

Korsgaard, Christine. 2009b. "Facing the Animal You See in the Mirror." *The Harvard Review of Philosophy* Vol. XVI: 2-7.

Korsgaard, Christine. 2011. "Interacting with Animals." In *The Oxford Handbook of Animal Ethics* (ed. Tom Beauchamp and R.G. Frey), Oxford University Press: 91-118.

Korsgaard, Christine. 2012a. "Getting Animals in View." *The Point Magazine*: 2-26.

Korsgaard, Christine. 2012b. "A Kantian Case for Animal Rights." *Animal Law* (ed. Margot Michel, Daniela Kühne, and Julia Hänni), St. Gallen: 3-27.

Korsgaard, Christine. 2013a. "Kantian Ethics, Animals, and the Law." *Oxford Journal of Legal Studies*: 1-20.

Korsgaard, Christine. 2013b. "Personhood, Animals, and the Law." *Think* 34, Vol. 12: 25-32.

Korsgaard, Christine. 2013c. "The Relational Nature of the Good." In *Oxford Studies in Metaethics* (ed. Russ Shafer-Landau), vol. 8, Oxford University Press: 1-26.

Korsgaard, Christine. 2014. "On Having a Good." *Philosophy* Vol. 89, Issue 3: 405-429.

Korsgaard, Christine. 2018. *Fellow Creatures*. Oxford University Press.

Nussbaum, Marha. 2023. *Justice for Animals*. Simon & Schuster.

Pears, David. 1984. *Motivated Irrationality*. Clarendon Press.

Pybus, Elizabeth and Broadie, Alexander. 1974. "Kant's Treatment of Animals." *Philosophy* Vol. 49: 375-383.

Rawls, John. 1971. *A Theory of Justice*. Belknap Press.

Raz, Joseph. 2008. "On the Guise of the Good." *Oxford Legal Studies Research Paper Series* 43.

Ridge, Michael. 2005. "Why Must we Treat Humanity with Respect?" *European Journal of Analytic Philosophy* Vol. 1, No. 1: 57-73.

Ripstein Arthur and Tenenbaum, Sergio. 2020. "Directionality and Virtuous Ends." In *Kant and Animals* (ed. John Callanan and Lucy Allais), Oxford University Press: 139-156.

Stocker, Michael. 1979. "Desiring the Bad." *The Journal of Philosophy* Vol. 76: 738-753,

Theunissen, L. Nandi. 2018. "Must We Be Just Plain Good?" *Ethics* Vol. 128: 346-372.

Timmermann, Jens. 2005. "When the Tail Wags the Dog." *Kantian Review* Vol. 10: 128-149.

Timmermann, Jens. 2006. "Value Without Regress." *European Journal of Philosophy* Vol. 14, Issue 1: 69-93.

Velleman, J. David. 1992. "The Guise of the Good." *Nous* Vol. 26.1: 3-26.

Velleman, J. David. 2000. *The Possibility of Practical Reason*. Oxford University Press.

Williams, Bernard. 1970. "The Self and the Future." *The Philosophical Review* Vol. 79, No. 2: 161-180.

Wood, Allen. 1999. *Kant's Ethical Thought*. Cambridge University Press.

Wood, Allen. 2008. *Kantian Ethics*. Cambridge University Press.

[1] MS, AA 06: 443.14-16.

[2] Anth, AA 07: 127.04-10.

[3] (Pybus and Broadie, 1974, p. 375), quoted in (Timmermann, 2005, p. 138) and (Birch, 2020, 211).

[4] (Bentham, 1907, 31111).

[5] In the words of Chignell, Korsgaard is "grasping for...the holy grail for Kantians...who are also animal advocates...to expand the *Formula of Humanity* into what might be called the *Formula of Sentient Animality*" (Chignell, 2020a, p. 1).

[6] GMS, AA 04: 429.10-13, emphases omitted.

[7] Evidence for Korsgaard's commitment to this premise may be found throughout her publications (Korsgaard, 1996a, p. 116; 1996b, p. 122; 2004, p. 93; 2009, p. 15; 2011, p. 106; 2012a, p. 11; 2012b, p. 7; 2013a, p. 11; 2018, preface). However, she sometimes uses slightly different terminology. For example, in *Fellow Creatures*, she writes of absolute goodness: "because we are rational, we cannot decide to pursue an end unless we take it to be absolutely good" (2018, section 8.4.1).

A complication arises here: it is not clear that Korsgaard is using these terms univocally. For example, Korsgaard explains absolute goodness in two different ways in *Fellow Creatures*: "I have characterized an absolute good as one that is good from every point of view, and also as one that can be included in a shared or common good which we can all pursue together" (2018, section 8.8.3). Not only is it unclear whether either of these definitions of absolute goodness is equivalent to the account of objective goodness, taken from Korsgaard's earlier work, that I give in the main text above (see note 9 and the sentence to which it is appended), but it also is unclear whether these two definitions of absolute goodness are equivalent to each other (e.g., there does not seem to be any reason why there could not be something that is good from every point of view but that could not be pursued together (one might think of Kant's joke about Francis I, "what my brother Charles wants (Milan), I want too" (KpV, AA 05: 28.22-23)).

[8] See (Kahn, 2013, section 2).

[9] (Korsgaard, 1996a, pp. 118, 258, 260, 262). Korsgaard also explains objective goodness in terms of justification and having sufficient reasons for the pursuit of an end (Korsgaard, 1996a, pp. 118-119; 2012b, pp. 7-8). It is on these grounds that Korsgaard argues that premise (1) articulates an internal, or constitutive, norm of rational action. The idea is that it is not possible to act without representing one's ends as objectively good (Korsgaard, 2009, p. 32). Note that Korsgaard is not the only philosopher, or even the only Kantian, committed to this thesis (see Allison, 1990, p. 91; Engstrom, 1992, p. 760; Herman, 1993, p. 217; and Wood, 1999, p. 129 or 2008, p. 91).

[10] (Korsgaard, 1996a, pp. 119, 121, 259).

[11] (Gewirth, 1978, p. 49); (Pears, 1984, chapters 9-10); and (Raz, 2008).

[12] (Stocker, 1979); (Velleman, 1992); and (Velleman, 2000, chapter 5).

[13] Godfrey-Smith rejects premise (1) on the grounds that he does not expect others not to interfere with the pursuit of his ends, much less to help with that pursuit:

> I hope for it, am pleased when it happens, and am also glad to live in a society where interference is discouraged. But I don't think that, just because another person will see that what I am doing makes sense for someone in a situation like mine, they will think they have reason to help me. (Godfrey-Smith, 2021, pp. 7-8)

Korsgaard does appeal to the idea that we expect others not to interfere with the pursuit of our ends in order to motivate premise 1 (2018, section 8.4.3). But there are several problems with Godfrey-Smith's argument. First, Godfrey-Smith commits the fallacy of denying the antecedent. If Korsgaard's idea about expecting non-interference is misguided, that does not impugn premise (1); it impugns this particular way of arguing for premise (1). Second, Godfrey-Smith mistakes the nature of the idea: it is supposed to have a *ceteris paribus* condition. That is, all else being equal, we do not expect others to interfere with the pursuit of our ends. The "all else" in this case includes: (i) other agents are behaving permissibly, and (ii) other agents do not have good reason to interfere with the pursuit of our ends.

Godfrey-Smith might reply that, (a) even with the *ceteris peribus* conditions factored in, he does not accept Korsgaard's idea about non-interference. Further, he might contend that (b) this non-interference idea is a presupposition of premise 1 and, thus, if it turns out to be mistaken, then premise (1) falls with it. But, this dual-reply does not work. Even if Godfrey-Smith could provide some convincing evidence that he does not accept Korsgaard's idea about non-interference, Korsgaard might note that this would show only that Godfrey-Smith is being irrational--or, more simply, that he has not realized all the implications of the way that he represents his ends (in much the same that he has not realized all the implications of the axioms of Peano arithmetic). I conclude that premise (1) can weather Godfrey-Smith's attack. More direct attacks on this premise may be found in (Hill, Jr., 2002, chapter 8) and (Kahn, 2013). Concerns about Korsgaard's ascription of this premise to Kant may be found in (Kerstein, 2006) and (Kahn, 2021, section 2.4).

[14] (Korsgaard, 1996a, p. 116; see also 2018, section 8.4.1).

[15] Nussbaum criticizes Korsgaard's antirealism on the grounds that (i) it does not comport with Korsgaard's goals and (ii) it is unnecessary:

> Korsgaard insists...that all value is a human creation. It does not exist "out there" to be discovered...Korsgaard's reasons for her view are Kantian: our reason is limited in scope, and does not entitle us to make claims that go outside the bounds of our experience...[but] this controversial metaphysical position is not necessary for [Korsgaard's] conclusion about the worth of animal lives, and is actually inappropriate if what we are pursuing is the creation of good political principles that can unite people of different religious and metaphysical views. If we are seeking political principles, as I think we both are, we must endeavor to construct a political and legal view that can ultimately be acceptable to people holding many different metaphysical and secular conceptions of the ultimate value. And this means that we must not attempt to justify a fully comprehensive ethical view. (Nussbaum, 2023, p. 72)

I would like to say four things in response to this.

First, I do not think that Korsgaard is engaged in the project of constructing a Rawlsian political and legal view. Let me explain. One of the objections that critics of Rawls raised against his original articulation of justice as fairness in *A Theory of Justice* was that it was based in a comprehensive ethical view. Rawls took this objection seriously, and over the next thirty years he sought to articulate what he called a political, rather than a comprehensive or metaphysical, liberalism--a version of justice as fairness that could be worked up from the general discourse (and generate an overlapping consensus) in any reasonable constitutional democracy. Nussbaum's own capabilities approach grew out of a different objection to Rawls, the objection that he was too preoccupied with his so-called primary goods and, in particular, income and wealth. But, Nussbaum, like Rawls, takes seriously the idea that liberalism should not be based in a comprehensive ethical view. The problem, however, is that Korsgaard is not engaged in the same project: Korsgaard, unlike Nussbaum, is not trying to articulate a theory of justice--she is (unapologetically) trying to articulate a comprehensive ethical view.

Second, even if Korsgaard were trying to articulate a theory of justice, the claim that such a theory must be political rather than comprehensive is contentious. Indeed, many criticize Rawls' later versions of justice as fairness on precisely these grounds, arguing that, on the one hand, a theory of justice grounded in a comprehensive ethical view could garner an overlapping consensus and, on the other hand, a theory of justice that is not grounded in a comprehensive ethical view runs the risk of devolving into relativism.

Third, as my reconstruction of Korsgaard's regress argument hopefully makes clear, her antirealism, encapsulated in premise (2) of my reconstruction, is, in fact, an essential ingredient: if antirealism is removed from the argument, then it falls apart. But, Korsgaard's conclusion about the worth of animal lives is based on an extension of the regress, so, if the regress falls, then this conclusion is no longer supported.

Fourth and finally, as the paragraph to which this note is appended hopefully makes clear, Korsgaard's antirealism is not based on any broad views about the nature of reason, its limits, or the bounds of experience (*pace* Nussbaum).

[16] (Korsgaard, 2004, p. 93; see also 2012b, p. 13; and 2018, section 1.3).

[17] (Korsgaard, 2018, section 8.3.2). The technical way that she expresses this in *Fellow Creatures* is to say that "all importance is tethered" (2018, section 1.3.2).

[18] (Korsgaard, 2013a, p. 15); see also (Korsgaard, 2013c; and 2014).

[19] In *Fellow Creatures*, Korsgaard accompanies this with an argument about how valuing "is originally an activity of life, a feature of a sentient creature's relationship to herself" (2018, section 8.3.3).

[20] (Korsgaard, 1996a, p. 345; see also pp. 260-261 and her 2018 sections 8.4.2-8.4.4). Korsgaard sometimes articulates premise (3) independently of premise (2), as in the following passage: "It is as if each of us said to herself, 'The things that matter to me are important, because I am important; what happens to me matters, because I do'" (Korsgaard, 2011, p. 106, repeated almost word-for-word in her 2012a, p. 11; 2012b, p. 9; and 2013, p. 14).

[21] There is some ambiguity here. In the block quote above, Korsgaard says that "the presence of a good will" is the unconditioned condition of the objective goodness of an agent's happiness. But, in some places, she seems to suggest that the unconditioned condition of the objective goodness of a choice is not the presence of a good will in general but, rather, an instance of good willing.

[22] As Korsgaard puts the point, using technical language, "Humanity, completed and perfected, becomes personality" (Korsgaard, 1996a, p. 114). Precisely this step seems suspect to many. Xs and perfect Xs do not in general have the same value, so why think rational nature has whatever value good wills have?

[23] One reason for being suspicious of this move in Korsgaard′s argument is that, even if we accept that she is right about perfectly rational beings, it raises deep questions about bounded rationality. That is, even if we grant that a perfectly rational being, in representing her ends as objectively good, presupposes that her capacity for rational choice is unconditionally good, it does not follow that an imperfectly rational being does so. An imperfectly rational being might have various beliefs that, although false, make it more rational for her to suppose that the end she is pursuing, rather than her capacity for rational choice, is unconditionally good. For example, a 12th century theist, steeped in Thomist metaphysics, very well might be more rational in supposing that her ends are unconditionally good in themselves rather than that her capacity for rational choice is so. And this is so even if we grant, for the sake of argument, that this step in Korsgaard′s derivation of the Formula of Humanity holds for perfectly rational beings. As far as I am aware, this line of criticism has not been developed in the (voluminous) literature on Korsgaard′s regress.

[24] See, in addition to the sources cited in previous notes, (Ridge, 2005), (Timmermann, 2006), (Kahn, 2017), (Bukoski, 2018), and (Theunissen, 2018). A novel rejection of Korsgaard′s premise (2), on cognitive grounds, may be found in Garthoff: ″If it is true both (i) that belief constitutively involves responsiveness to reasons and (ii) that belief does *not* constitutively involve representation of reasons, then it cannot be, as Korsgaard′s constructivism demands, that all reasons are constructed through the activity of critical reason″ (2020, p. 208).

[25] See (Korsgaard, 2004, pp. 85-86; 2006, p. 5; 2009, pp. 5, 15; 2011, p. 103; 2012a, pp. 16-17; 2012b, pp. 8-9; 2013a, pp. 9-10; and 2013b, p. 26).

[26] (Korsgaard, 2008, p. 5). Korsgaard′s account of the difference between humans and other animals in *Fellow Creatures* is more expansive. In sections 2.4.1, 3.2.2, 3.3.3, 3.4.1, and 12.2.1, the difference is explained in terms of self-governance, self-awareness, and self-constitution, much as in her earlier work. In addition, in section 3.2.3 Korsgaard claims that humans choose action-purpose pairs whereas other animals choose merely actions; in section 3.4.2 she argues that humans, unlike other animals, identify with our species as a group and, as such, humans have collective agency; and in section 3.5.2 she argues that other animals ″cannot view the world scientifically″ because ″they have no conception of a world whose operations have nothing to do with them.″

[27] (Korsgaard, 2011, p. 108).

[28] (Korsgaard, 2011, p. 108; 2018, section 8.5.4).

[29] (Korsgaard, 2018, section 8.5.5).

[30] The stronger version is in (Korsgaard, 2004, pp. 104 and 196; 2009a, p. 15; 2012a, pp. 22-23; 2012b, p. 13; and 2013, p. 14). The weaker version is in (2009b, p. 5). In some places, Korsgaard seems to begin with the weaker version but then to move to the stronger one (2008, p. 5; and 2018 sections 8.5.3 (weaker) and 8.5.4 (stronger)).

[31] (Kahn, 2019, 199n21); (Birch, 2020, pp. 5-8); and (Chignell, 2020a, p. 10; and 2020b, p. 214). None of the respondents, however, considers Korsgaard′s reply to them.

[32] (Korsgaard, 2004, 104n66).

[33] (Korsgaard, 2012b, p. 13; 2018, section 8.5.4).

[34] (Korsgaard, 2012b, pp. 13-14; see also her 2013a, pp. 14-15 and 2018, section 8.5.4).

[35] (Williams, 1970).

[36] Consider the hew and cry subsequent to the public disclosure about cadavers being used as crash test dummies in the 1980s (Bergelson, 2009, pp. 177-178).

[37] (Korsgaard 1996, p. 123; also pp. 125 and 241).

[38] (Denis, 2000), (Ripstein and Tenenbaum, 2020), (Timmermann, 2005).