

Meriting Concern and Meriting Respect

Jon Garthoff

RECENTLY THERE HAS BEEN A SPATE of interest, among Kantian moral theorists, in the moral standing of conscious animals.¹ This interest is somewhat surprising, since Kant's approach to moral theory appears uncongenial to obligations toward animals. Kant's theory traces human moral standing to the possession of a cluster of rational capacities, including cognitive capacities to represent reasons as reasons and to conceive of oneself as existing over extended periods of time. This emphasis makes Kantian theory famously effective at distinguishing the moral status of human persons from that of non-rational animals – in part by providing resources for explaining the wrongness of coercion and deception – but it also makes the theory famously ineffective at accounting for whatever moral status human persons share with non-rational animals.

Not less surprising than this Kantian interest in the moral standing of animals is the optimism Kantians exhibit for the prospect of incorporating animal moral standing into Kantian moral theory without contorting its other attractive features. While everyone acknowledges amendments to Kant's theory are needed, leading theorists maintain that these amendments need not be highly extensive. They contend in particular that it is possible to incorporate animal moral standing into Kantian theory without abandoning the theory's *logocentrism*. As I understand it, logocentrism is the thesis that rationality is the only thing that is valuable independent of the value of anything else.² This claim is typically conjoined, moreover, to the additional claim that everything else of value depends for its value on its relationship to rationality.

Logocentric Kantians who seek to vindicate animal moral standing must avoid drawing what Allen Wood has called Kant's "ghastly inference."³ This inference appears in Kant's *Conjectures on the Beginning of Human History*; there he writes that when a human first

... said to the sheep, "nature gave the skin you wear not for you but for me," and then took it off the sheep and put it on himself (*Genesis* 3:21), he became aware of the prerogative he had by nature over all animals, which he no longer saw as fellow creatures, but as means and tools at the disposal of his will for the attainment of the aims at his discretion.⁴

¹ See, for example, Allen W. Wood (1998), Onora O'Neill (1998), Christine Korsgaard (2004, 2007) and Rae Langton (2007).

² Wood glosses "logocentrism" somewhat differently, calling it "the idea that rational nature, and it alone, has absolute and unconditional value." See Wood (1998), 189. I use the formulation in the text to avoid disputes about the notions of "absolute" and "unconditional." For more on those notions, see Korsgaard (1983, 1986b), Karl Ameriks (1989), chapter four of Wood (1999) and Langton (2007).

³ See Wood (1998), 190; see also Immanuel Kant (1786), 8:114, and Kant (1798), 7:127.

⁴ Kant (1786), 8:114.

The ghastly inference is to conclude, from the logocentrism that vindicates the superior moral standing of humans, that animals are mere things to be used in the same manner as resources like timber and coal. The task for logocentric Kantians is thus to show that their logocentrism need not have this implication. To accomplish this, they must resist a simple answer to the question of whether non-rational animals have moral standing. In order to avoid the ghastly inference, they assert that animals have standing – but they also assert that animals have this standing only by virtue of their relationship to rationality, a capacity which by hypothesis they do not possess.

In this essay I raise doubts about the prospects for accommodating the moral standing of animals within logocentric Kantianism, and I argue as a consequence that more radical changes should be made to Kant's theory. I share the motivations behind recent efforts to accommodate animal moral standing within this theory: I believe both that many animals have moral standing and that Kantianism is the most promising approach to moral theory. I argue here, however, that the best way to incorporate animal moral standing into Kantianism is to relax its logocentrism and maintain that consciousness – understood as a non-rational capacity – is a locus of moral standing independent from its relationship to rationality. I do so by arguing for a position more typically asserted by utilitarians: that consciousness, not rationality, is the capacity in virtue of which there are both reasons to promote an individual's well-being and obligations not to violate or denigrate that individual. As I motivate in this essay, however, this non-Kantian criterion of moral standing is compatible with a Kantian account of why rational individuals have higher moral standing than non-rational individuals. I thus begin to make the case for both a distinctive account of the moral standing of animals and a new understanding of the moral standing of persons.

1. Wood's Logocentrism

Before I present my own proposal, I first briefly consider the approach to these issues taken by two logocentric Kantians, Allen Wood and Christine Korsgaard. I begin with Wood's approach.⁵ Wood observes that many conscious but non-rational animals have capacities that are "recognizable fragments" of rationality. These include the capacities for desire and preference, which lead animals to experience pain and frustration when their desires and preferences are not fulfilled. Wood suggests that respect for rationality itself demands we value these fragments of rationality by understanding them as the "infrastructure ... of rational nature."⁶ While having the capacities for desire and preference is not sufficient for having rationality, Wood's thought is that these capacities are so intimately related to rationality that respect for

⁵ See Wood (1998).

⁶ Wood (1998), 200.

rationality can obligate us to value animals which have them. On Wood's view, animals with these capacities have moral standing on their own, by virtue of these quasi-rational capacities; yet they have this standing derivatively, by virtue of the relation between these capacities and full-blown rationality.⁷ To capture this complex status, Wood suggests we say we have obligations "in regard to" animals but that we have obligations "toward" only rational individuals.

Wood's position is that there is no need to abandon Kant's logocentrism, so long as we abandon Kant's focus on individuals who possess rational capacities in the full-blown sense in which fully developed humans possess them. In keeping with a standard philosophical use, we may follow Wood in referring to individuals who have these full-blown rational capacities as "persons." Wood suggests that we amend Kant's theory to shift focus away from persons without amending the theory to shift focus away from the value of their definitive capacity.

His proposal, more specifically, is that we reformulate Kant's canonical statement of the content of his moral theory: his formula of humanity. Kant expresses this principle in section II of the *Groundwork of the Metaphysics of Morals*:

So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.⁸

To understand what this principle says, it is crucial to note that "humanity" is a technical term in Kant. The English term "humanity" typically refers either to membership in the human species or to benevolent dispositions toward others. But for Kant, "humanity" refers instead to the rational capacities of embodied individuals.⁹ The formula of humanity thus states a requirement on how individuals with these rational capacities are to be treated; it says nothing directly about species-membership or benevolent dispositions. This much is familiar to students of Kant. But less commonly noticed, as Wood observes, is that this expression of the formula of humanity limits the scope of morality to cases where humanity is instantiated in persons: it is in your own *person*, or in the *person* of any other, that humanity cannot be treated as a mere means. Thus on Kant's formulation, only full-blown rationality – personhood – generates obligations.

⁷ See Wood (1998), 200.

⁸ Kant (1785), 4:429.

⁹ Wood glosses humanity as "the capacity to set ends according to reason"; see Wood (1998), 189. I use the formulation in the text to avoid disputes about the notion of end-setting and to use less distinctively Kantian lingo, since Kant's claims about humanity resonate beyond his own approach. A full treatment of this issue would explore the relations between humanity and the capacities to obey the moral law, to value things, and to adopt and pursue aims; this full treatment would engage not only *Groundwork* but also other Kantian texts, including *Religion Within the Boundaries of Mere Reason*. See Kant (1793), especially Book One.

Wood refers to this as Kant's "personification principle." By hypothesis non-rational animals are not persons, so, if the formula of humanity as Kant states it is an exhaustive account of obligation, then there can be no obligations in regard to animals.¹⁰ Wood recommends, in an effort to preserve logocentrism in the face of this entailment, that we remove the personification principle from the formula of humanity. He thus recommends recasting this principle as:

So act that you use humanity always at the same as an end, never merely as a means.

Once we remove the personification principle from the formula of humanity, Wood maintains, it is possible for the formula of humanity to account for obligations in regard to animals. Although animals do not possess humanity, Wood claims, their quasi-rational capacities must be respected if we are to treat humanity as an end in itself; a failure to respect quasi-rational capacities would express contempt, on his view, for the value of rationality itself. Wood claims this view is in keeping with the spirit of Kant's theory,¹¹ and as he notes it is attractive independent of worries about the standing of animals. There are longstanding concerns, in view of Kant's logocentrism, about the ability of Kantian moral theory to incorporate obligations in regard to humans who are not rational in a full-blown sense, like infants, the severely mentally disabled and the temporarily incapacitated.¹² Dropping the personification principle from the formula of humanity removes an important barrier to a logocentric explanation of the moral standing of these individuals.

Wood does not flesh out his strategy for defending logocentrism in detail. A schematic worry applies to any attempt to defend this strategy, however, regardless of how it is developed. If the moral standing of animals results from the relation between their capacities and rationality, this worry proceeds; then it is unclear why we must value their quasi-rational capacities when these fail to constitute the infrastructure of rationality. If the only reason we are obligated to treat animal capacities as valuable is that they constitute infrastructure of rationality, that is, then it appears we are obligated to treat them as valuable only when they manifest in rational beings.¹³ Compare:

¹⁰ See Wood (1998), 193-194. If nonhuman animals possess the relevant rational capacities – the great apes, say – then these animals are persons, and so are already accorded moral standing by the formula of humanity. For more on the moral standing of great apes, see Paolo Cavalieri and Peter Singer (1993).

¹¹ Wood believes Kant is committed to relaxing the personification principle by virtue of his view that non-rational humans have moral standing and his arguments that we must treat animals well in order to perfect ourselves. See Wood (1998), 198-202. Wood argues this case in part in anticipation of the objection that dropping the personification principle is an ad hoc solution to a deep problem with Kant's theory.

¹² See Wood (1998), 198-199.

¹³ Wood considers this worry and denies that it has force; see Wood (2008), 101-105. As I indicate in the text, however, his writings leave me unclear on the details of his response to it.

if we value the interstate system in the United States because it enables efficient travel, then we are committed to valuing bridges, since they are an essential constituent of the interstate system by virtue of which it enhances efficiency. But this does not commit us to valuing bridges wherever they may be found; in particular, bridges not reachable by road from the interstate need not be valued, as they may not enhance efficiency. Or consider that embodiment may also be an essential constituent of human rationality. This fact, if it is a fact, may commit us to valuing our embodiment insofar as it is an essential constituent of rationality, but would not thereby commit us to valuing the embodiment of mountains, trees or stuffed animals.

This worry brings out a danger of Wood's strategy – namely that it risks recapitulating a mistake in Kant's own account of obligations in regard to animals. As Wood explains, Kant's view is that these obligations derive from the corrupting influence that mistreatment of animals has on human character; Kant thus characterizes obligations concerning the treatment of animals as obligations of self-perfection held toward ourselves.¹⁴ This is the only way he can vindicate the existence of such obligations, since he attributes no standing to animals themselves. Wood's approach avoids this mistake, instead properly tracing these obligations to a moral status animals have by virtue of their own capacities. This strategy still makes their standing dependent on their relationship to rational capacities, however, and so still makes their standing derivative from that of rational individuals. This is not a ghastly inference, but it is a worrying one.

This section's discussion is meant not to rebut Wood's proposal, but only to call attention to its outstanding problems. Wood does not intend his proposal to entail that possession of any essential condition for rationality suffices for moral standing. As I explain below, moreover, I follow Wood in dropping the personification principle from the formula of humanity, though I believe that we should replace it with a new criterion of moral standing – consciousness – rather than drop it altogether. The present worry is intended to expose only that more work needs to be done to flesh out Wood's account; until further details are provided, we cannot have confidence that what we ordinarily take to be our obligations toward animals coheres well with the infrastructure of rationality hypothesis of their origin, or that this hypothesis coheres well with other commitments of Kant's moral theory.

2. Korsgaard's Logocentrism

Christine Korsgaard offers an alternative strategy for incorporating obligations in regard to animals into Kantian moral theory.¹⁵ Although this strategy differs from Wood's, it shares his ambition of preserving Kant's logocentrism where this is the claim that rationality is the only thing valuable inde-

¹⁴ See Wood (1998), 191-195; see also Kant (1797).

¹⁵ See Korsgaard (2004, 2007).

pendent of the value of anything else. As a prelude to discussing Korsgaard's account of animal moral standing, it will help to first consider her argument for human moral standing, since her case for the former is an extension of the latter.

Korsgaard's defense of the formula of humanity is inspired by a passage from Kant's *Groundwork*; it is known as her regress argument, and versions of it appear in a variety of locations in her work.¹⁶ I discuss this argument in section 3 below, where I explain both why I do not believe the argument succeeds as Korsgaard understands it and why I believe that an analog to the argument does succeed. For the moment I provide only a quick gloss of the argument, so that I can present Korsgaard's strategy for extending it to the case of animals.

The argument begins by claiming that, when a person reflectively endorses the fulfillment of his desires, he is committed to the objective value of this desire-fulfillment; this commitment entails both that he is rational to pursue the fulfillment of these desires and that he can justifiably complain when others interfere with his pursuit of this. But since we know that some desire-fulfillment is not objectively good, Korsgaard argues, the fact that a person desires something is not a complete explanation of the value of the desire being fulfilled. Korsgaard argues further that only appeal to a thing valuable of itself, to a thing with no external conditions on its value, can fill this explanatory gap; appeal to any other sort of thing, she claims, would invite a further demand for explanation. On Korsgaard's view, the only thing valuable in this way is rational willing, so she concludes that whenever we act to fulfill a desire we are committed to the claim that rational willing confers value on its fulfillment. Since pursuing any desire commits a person to the claim that rational willing has this value-conferring status, she concludes further that every person is committed to treating rational willing as having this status, which is to say that every person is committed not to violate Kant's formula of humanity as she understands it.¹⁷

Korsgaard's regress argument relies on what she takes to be an inescapable feature of the psychology of embodied rational individuals – that when a person reflectively endorses a thing, he is committed to the objective value of that thing. In the context of a person's acting on a desire, Korsgaard argues that his endorsement of fulfilling the desire commits him to valuing rationality as having conferred value on this fulfillment. Her extension of this argument to obligations with regard to animals applies the same premise in the case where what a person endorses is a good of consciousness, such as experiencing pleasure (or avoiding pain). According to Korsgaard, all persons endorse the experience of pleasure (and the avoidance of pain) as a component of their good; and she claims further that all persons are thereby com-

¹⁶ See Kant (1785), 4:428-429; and see Korsgaard (1983, 1986a, 1986b, 1996a, 1996b, 2003).

¹⁷ For critical commentary on Korsgaard's regress argument, see Berys Gaut (1997), Michael Ridge (2005), William J. FitzPatrick (2005) and Jon Garthoff (2010b).

mitted to the value of pleasure (and the disvalue of pain) even when experienced by a non-rational animal.¹⁸

Unlike Wood, then, Korsgaard's view is not that we are committed to valuing capacities like desire, preference and consciousness because they help to constitute the infrastructure of rationality. Her view is, rather, that since we are animals we cannot but endorse the value of goods like pleasure that are enabled by these capacities of consciousness, and so we confer value on the goods of consciousness regardless of who enjoys them. Non-rational animals are unable to confer value on their own pleasure, although they shape their behavior in response to it. But since we are rational and we share in the good of pleasure, we confer value on pleasure and other goods of consciousness when these are enjoyed by animals. Korsgaard's ultimate conclusion is thus in one respect stronger, and in another respect weaker, than that which Wood purports to establish. It is stronger because it entails that the capacities of consciousness must be valued as such, while Wood purports to establish only that these capacities are of value because they help constitute the infrastructure of rationality. Korsgaard's conclusion is weaker, however, inasmuch as it claims that moral standing is conferred on consciousness through the exercise of rational capacities, while Wood claims that possessing the infrastructure of rationality constitutes moral standing in its own right, independent of the exercise of rational capacities.

Korsgaard seeks to naturalize value by locating the source of all value in the psychology of humans, but she also seeks to establish that a commitment to the value of both rationality and consciousness is involved in every exercise of rational agency by a conscious individual. The crucial premise for her extension of the regress argument to include the case of animals is thus that it is a necessary fact about all conscious rational individuals that, on reflection, they endorse the value of the goods of consciousness. Her explicit argument for this premise, however, is brief; in her principal essay on animals, it appears only in a footnote. In that footnote she takes up an objection similar to the worry raised about Wood's account in section 1 above: this is the objection that her argument, even if successful, establishes only that we are committed to valuing the goods of consciousness when they are experienced by rational individuals:

I think that the correct reply is a fairly familiar one – that anyone who made such a claim would be lying or engaged in self-deception. For comparison, imagine a white male who claims that in valuing his own freedom he is only valuing the freedom of white males: if, unknown to himself, he turned out to be a black woman (imagine a genetic test with somewhat startling results) then he would agree that his freedom doesn't matter. Our response would be that he's either insincere or deceiving himself, that he's suffering from a failure of reflective imagination. This kind of response is harder to articulate in talking about human beings and the other animals. We have to say your legislation against being tortured or hunted or eaten would stand even if you were not a rational being. And that claim is ambiguous: in one sense it would not, since you would

¹⁸ See Korsgaard (2004), especially 26-33, and see also Korsgaard (2009).

then lack the power to legislate. But that sense is irrelevant. I want to say: the content of the legislation would stand, even though its form would fall. Allan Gibbard helpfully proposed to me that I should make the point R. M. Hare's way: ask the challenger to imagine that he is about to be deprived of his rational nature, but may now settle the question whether he will afterwards be tortured or not. Can he really say: "in that case it won't matter"?¹⁹

I am not sure whether Korsgaard's response is adequate in the case of our commitment to value rationality; it seems at least speculative to assert that no person could have the attitudes she ascribes to her white male. Of greater significance to this essay, however, is that the case for universal reflective endorsement of the value of the goods of consciousness is, as Korsgaard acknowledges, yet more speculative. Rich traditions of valuation – found in Buddhism, Stoicism, Neoplatonism, Christianity and elsewhere – emphasize self-denial and the elevation of rarefied rational goods above relatively base sensual goods. A participant in such a tradition, it seems, might fail to see the value in pleasure or the disvalue in pain altogether. While this is not on its own a decisive response to Korsgaard, it does indicate that it would be a strength of an account of animal moral standing if it could do without such a controversial yet necessary premise.

3. The Formula of Humanity

I am thus not convinced by Korsgaard's argument that the formula of humanity, when properly interpreted, yields obligations in regard to animals. As background for the alternative strategy for grounding animal moral standing I present below, however, it will help to discuss Korsgaard's understanding of the formula of humanity further. In some crucial respects I depart from her understanding of both this principle and its justification, and recasting the principle helps point us toward a different principle which grounds animal moral standing. I here have space only to summarize the rationale for the departures from Korsgaard's position I recommend; I provide a fuller explication and defense of these departures elsewhere.²⁰

As I indicated earlier, I believe Korsgaard's regress argument fails as she understands it, but that an analog to this argument succeeds in establishing that all persons are obligated not to violate Kant's formula of humanity. I want now to sketch this analog, as a preliminary to making my case for obligations toward conscious animals. I do so by distilling those elements of Korsgaard's regress argument that I believe withstand objection and distinguishing these from those elements that I believe should be abandoned. As this is my aim, I couch the claims in my reconstruction of Korsgaard's argument in a way that is neutral between her constructivism and a realism about value. The idea is to present an argument true to Korsgaard's intent – formu-

¹⁹ Korsgaard (2004), 31-32. See also R. M. Hare (1963), 222-224.

²⁰ See especially Garthoff (2010b).

lated such that she or another Kantian constructivist could endorse it – but which admits of an alternative, realist interpretation. My reconstruction of her argument is thus an archetype, for which her regress argument and my realist interpretation are instantiations. The central claims of this archetype are as follows:

- (1) Some aims that are rational to pursue – paradigms include projects and relationships, such as careers and friendships – are finally valuable for a person to achieve just in case she treats them as finally valuable in action.
- (2) If some aims that are rational to pursue are finally valuable for a person to achieve just in case she treats them as finally valuable in action, then a complete explanation of the rationality of her pursuing these aims must invoke the claim that the exercise of her rational capacities helps sustain their final value.
- (3) If a complete explanation of the rationality of a person’s pursuing some aims must invoke the claim that the exercise of her rational capacities helps sustain their final value, then a complete explanation of the rationality of her pursuing these aims entails that her possession of rational capacities entitles her to be treated only in ways that are consonant with recognition of the fact that their exercise can sustain the final value of aims; let us say anything entitled to this sort of treatment *merits respect*.
- (4) If a complete explanation of the rationality of a person’s pursuing some aims entails that she merits respect by virtue of her possession of rational capacities, then a complete explanation of the rationality of her pursuing these aims entails that any individual with rational capacities merits respect.
- (5) If a complete explanation of the rationality of a person’s pursuing some aims entails that any individual with rational capacities merits respect, then all persons are obligated not to violate the formula of humanity.
- (6) A complete explanation of the rationality of a person’s pursuing some aims entails that all persons are obligated not to violate the formula of humanity.

I believe claims (1) through (6) can all be vindicated. But before I defend a realist interpretation of these claims, I want first to explain briefly why I reject Korsgaard’s version of the argument: her regress argument.

Korsgaard has ambitions beyond those of the realist interpretation of the argument, one of which is to rebut skeptical challenges to the formula of humanity. Her regress is so called because it shares its structure with putative cosmological proofs of the existence of God. These arguments posit a constraint on a complete explanation of causation, traditionally known as the principle of sufficient reason, and seek to show that this constraint can be satisfied only on the presupposition that God exists. Korsgaard is clear that this is the structure of her regress argument: she uses the term “unconditioned condition,” alluding to the canonical expression “unmoved mover,” and she explicitly analogizes her argument to pre-Kantian rationalist argu-

ments for the existence of a first cause.²¹ As with other applications of the principle of sufficient reason, the regress argument presupposes that a philosophical account of a fact is incomplete unless it rules out the possibility of that fact's failing to obtain. More specifically, Korsgaard claims that an account of the fact that a thing is valuable is incomplete if it fails to preclude all conditions that would defeat the claim that it is valuable. By contrast, a complete explanation precludes all such defeating conditions, including any putative defeating conditions which might be advanced as skeptical hypotheses.

To rebut skeptical challenges by satisfying the principle of sufficient reason, a theory of causes must posit a thing that both is the cause of all other things – including the cause of other things' causal powers – and is itself uncaused or self-caused; this thing is thus characterized as an unmoved mover. Similarly, for a thing to play the role in the regress argument that is analogous to that of an unmoved mover in a cosmological proof, it must be a condition on the value of all other things and must also itself be unconditionally valuable. Elsewhere I argue there is no unconditioned condition of all value in this sense and, consequently, that Korsgaard's regress does not succeed.²² This leads me to try to reconstruct the argument in a different way.

While I reject Korsgaard's regress argument, I agree with her that claims (1) through (6) can all be vindicated. In other work I argue for this in detail; I aim here only to summarize that argument, so that in section 4 I can explain how an analogous argument supports obligations toward conscious animals.²³ I begin with claim (1):

- (1) Some aims that are rational to pursue – paradigms include projects and relationships, such as careers and friendships – are finally valuable for a person to achieve just in case she treats them as finally valuable in action.

I start by noting that we do not have, for each of our finally valuable aims, an account of why it is rational to treat that aim as finally valuable that is independent of its being adopted as an aim. I note, that is, that we lack a wholly agent-independent explanation of the value of paradigmatic finally valuable aims, including interpersonal relationships (such as friendships and romantic relationships) and long-term projects (such as careers and hobbies). The following illustrates this claim: my vocation as a philosophy professor and my avocation as a Scrabble enthusiast generate reasons for action. By contrast the career as a politician that I might have had, but do not have, and the hobby as a chess player I might have had, but do not have, fail to gener-

²¹ Korsgaard (1983), 259. Like canonical users of the principle of sufficient reason, Korsgaard intends her regress argument to be purely *apriori* justified. This is among the ambitions that I abandon here.

²² See Garthoff (2010b).

²³ For much fuller defense of the central claims of this argument, see Garthoff (2010a, 2010b).

ate reasons for action.²⁴ Furthermore, I have reasons to adopt sub-aims of my actual aims, such as writing essays and delivering lectures – but there are no reasons for me to adopt sub-aims, such as raising funds and campaigning for office, of the aim I do not have of succeeding as a politician.

One might think I am rational to attend to my actual aims in these usual ways because being a philosophy professor is more valuable for me than being a politician and because being a Scrabble player is more valuable for me than being a chess player. Even if true, however, these claims play no role in explaining what I have reason to do when I deliberate among actions. To see why, suppose I suspect I should change my career or hobby. This entails that I have reason to deliberate about a possible change of career or hobby, but it does not entail that my actual aims no longer provide me with reasons for action. It may be rational not to deliberate about these matters, or to postpone this deliberation. If so, then it is rational to continue to treat my actual career and hobby as generating reasons for action – assuming these aims are in fact valuable for me to pursue, even if they are not necessarily the *most* valuable – just as I did before I suspected a change might be in order.

Or suppose I realize I have greater aptitude as a politician than as a philosophy professor, or that I live in a society that values politics more highly than philosophy.²⁵ Even then my actual aim determines what I have reason to do, in the context of deliberation among actions, provided it is sufficiently valuable. To explain this, we need to look beyond whatever value my aims have independent of my adopting them – their “choiceworthiness” – and advert to a fact about how I exercise my rational agency.²⁶ In order to explain why it is rational for me to try to succeed as a philosophy professor, but not to try to succeed as a politician, I must appeal to the facts that I aim to succeed as a philosophy professor and that I do not aim to succeed as a politician.

These considerations show that the reasons I have for performing this or that action are sensitive to the projects and relationships I have actually adopted; but they do not yet show that my adoption and pursuit of aims – the exercise of my rational capacities – helps sustain their final value, as follows from the conjunction of claim (1) with claim (2):

(2) If some aims that are rational to pursue are finally valuable for a person to achieve just in case she treats them as finally valuable in action, then a complete explanation of the rationality of her pursuing these aims must invoke the claim that the exercise of her rational capacities helps sustain their final value.

²⁴ To be more precise: the only sorts of reasons they provide are reasons to stop deliberating among actions and to begin deliberating among aims.

²⁵ Note that it might, under these circumstances, be rational for me to abandon my aim of succeeding at philosophy and adopt the aim of succeeding at politics, but that is a different claim.

²⁶ I adapt this notion of choiceworthiness from T. M. Scanlon; see Scanlon (1998), 131.

Earlier I rejected Korsgaard's supposition that an explanation is complete only if it satisfies the principle of sufficient reason, and with it her constructivist interpretation of (2). I suggest instead we regard claim (2) as supported by an inference to the best explanation.²⁷

I suggest that the best explanation of the fact that some aims are finally valuable just in case they have been adopted is that the adoption of aims helps sustain their value. More specifically, I suggest in what follows that the hypothesis that the exercise of rational capacities can sustain the final value of aims is the best explanation of the conjunction of (i) the fact that it is rational in action to treat actually adopted choiceworthy aims, but not non-adopted choiceworthy candidate aims, as valuable and (ii) the fact that it is rational to treat idiosyncratic features of choiceworthy aims as valuable even when these features do not contribute to these aims' choiceworthiness. The hypothesis that my adoption of Scrabble as a hobby helps sustain the final value of my success in Scrabble, for example, is the best explanation of why it is rational for me to treat this success as finally valuable but not to treat my success in chess as finally valuable. And this hypothesis is the best explanation of why it is rational for me to treat idiosyncratic constituents of Scrabble success as non-instrumentally valuable.

One might think against this hypothesis that general features of human practical reason fully explain the rationality of treating only actually adopted aims as valuable in action. Humans are finite, and so they cannot entertain every consideration that bears on what they should do; nor are they able to pursue all the aims that are worthwhile for them to pursue. They must organize the exercise of their agency by adopting aims, and they must order their pursuit of these aims both temporally and hierarchically.²⁸ One might think that candidate aims have whatever value they have, and that it is only a consequence of my finitude that when I act I am rational to treat some candidate aims as valuable (the choiceworthy ones I have adopted) but not others (those I have not adopted or are not choiceworthy).

But this is not right. To see why, we should focus not on features of our aims that make them worthy of choice, but instead on their idiosyncratic features that do not contribute to their choiceworthiness. My pursuit of Scrabble as a hobby, for example, is choiceworthy largely because Scrabble expertise consists in mental skills of wide application, like facility recalling information from memory, large vocabulary and quickness in the performance of arithmetic. But Scrabble expertise also consists in part in arcane skills that are far less useful, such as knowing all the words of English that contain a "Q"

²⁷ Ridge (2005) labels Korsgaard's own argument as an inference to the best explanation, but I believe this is misleading. For while it is true that deployments of the principle of sufficient reason appeal to standards of explanation, the label "inference to the best explanation" is commonly reserved for arguments which deny the high explanatory standards of pre-Kantian rationalism.

²⁸ For extensive developments of this claim, see Part III of Rawls (1971); see also Henry Richardson (1997) and Michael Bratman (2007).

but not a “U” and having familiarity with the details of the *Official Scrabble Player’s Dictionary*. If the value for me of pursuing Scrabble is exhausted by whatever value this aim has independent of my adoption of it, then all that I should value about my hobby are the features that make it choiceworthy, such as that it enables me to develop and exercise my cognitive capacities. In particular, I should not treat idiosyncrasies of Scrabble as helping constitute the hobby’s final value, and this should lead me to regard peculiarities of pursuing Scrabble as impurities or inefficiencies of the hobby as a way to develop and exercise cognitive capacities.

But this need not be the case. For the Scrabble enthusiast, the game’s idiosyncrasies are, in the context of deliberations about what to do in pursuing Scrabble, on a par with features that make it choiceworthy; *and this does not appear, on reflection, to be a mistake*. If the entire activity of Scrabble-playing consisted in deploying arcane knowledge, Scrabble would be a relatively unchoiceworthy aim. But since Scrabble expertise is, we may stipulate, a choiceworthy aim for the enthusiast, the peculiar Scrabble-reasons have no second-class status. There is nothing irrational about valuing success in the idiosyncratic features of our choiceworthy aims.

The best explanation of this is that Scrabble and other choiceworthy aims have a mode of value sustained by their adoption. Scrabble is choiceworthy – there is reason for me to adopt it as an aim – because it constitutes the exercise of cognitive capacities and is a means to improving these capacities. But once I adopt Scrabble as a hobby, it can also be rational for me to treat success in Scrabble as valuable for its own sake. That is why success in Scrabble’s idiosyncratic features is valuable for me. It is thus rational for enthusiasts to treat playing the game as valuable for its own sake, but it is not rational for others to treat this aim as valuable in that way. The only relevant point of difference between these two groups of individuals is that the former has adopted the aim and the latter has not and, as a consequence, the best explanation of these facts is that actual adoption of an aim can sustain a value – its final value – that it would not otherwise have.

Thus far I have used a hobby to illustrate that many aims have their final value sustained by the exercise of rational capacities. But if the preceding considerations motivate the claim that the exercise of rational capacities can sustain the final value of hobbies, then they also motivate the claim that the exercise of rational capacities can sustain the final value of other projects. Arcane insect knowledge is valuable for an entomologist, but not for others. The claim also applies to relationships. It is rational to value actual friendships and romantic relationships for their own sake; it is a gross error to value a particular friendship only as a means to happiness or as constituting a way of having *a* friend. This is why it can be rational to treat the idiosyncratic constituents of a friendship as valuable rather than as drags on its ability to constitute the more general value of having a friendship. My suggestion is

that these facts are best explained by the hypothesis that the actual adoption of a relationship as an aim helps sustain its final value.²⁹

This claim is in some respects similar to Korsgaard's claim that the exercise of rational capacities is the source of all value, but it is much weaker.³⁰ First, it does not entail that *all* value is sustained by the exercise of rational capacities; indeed, the argument for the claim presupposes some value is *not* sustained in this way, since it deploys a notion of choiceworthiness that applies to candidate aims independent of whether they are adopted. Second, my claim does not entail that rational capacities can generate value from nothing; I claim that only *choiceworthy* aims can have final value when they are adopted. Independent of my adoption of it as an aim, Scrabble is non-finally valuable as a way to constitute both a hobby and the exercise of cognitive capacities. Independent of my adoption of it as a career, being a philosophy professor is non-finally valuable as a way to constitute both a career and the exercise of cognitive capacities. Independent of my actually developing relationships with my friends, candidate friendships are non-finally valuable as a way to constitute a friendship and also as a way to exercise emotional capacities. When I adopt these aims, however, I claim that they acquire a mode of value – their final value – that they did not previously have. Since this way of vindicating (2) avoids the surprising claim that all value is conferred by rational choice, it provides a weaker and more plausible basis from which to argue for the formula of humanity.

Thus far I have argued that a person's exercise of rational capacities can sustain the final value of her choiceworthy aims. If we accept this claim, the reconstructed argument continues with claim (3), then we must regard her as worthy of respect:

- (3) If a complete explanation of the rationality of a person's pursuing some aims must invoke the claim that the exercise of her rational capacities helps sustain their final value, then a complete explanation of the rationality of her pursuing these aims entails that her possession of rational capacities entitles her to be treated only in ways that are

²⁹ Note that this is *not* an endorsement of the view, advanced by David Schmitz and Harry Frankfurt, that appeal to instrumental reasons for treating an aim as finally valuable can fully rationalize treating the aim this way. See Schmitz (1994) and Frankfurt (2004), especially section 10 of chapter two. I cannot argue fully against this view here, but I would briefly mention a few reasons for dissent: valuing an aim finally for instrumental reasons is unstable, since reflecting on how one came to adopt the aim undermines the conviction that it is finally valuable; valuing an aim finally for instrumental reasons makes the generation of final value too volitional, since it opens widely the conditions under which persons can sustain the final value of aims; and this model assimilates paradigmatic cases of rationality, such as treating choiceworthy projects and relationships as finally valuable, to cases of borderline irrationality, such as valuing a higher power for its own sake in order to stay sober or valuing a diet for its own sake in order to lose weight. For much fuller development of this argument, see Garthoff (2010a).

³⁰ See Korsgaard (1983, 1986a, 1986b, 1996a, 1996b, 2003). Korsgaard's position is inspired by Kant; see especially Kant (1785), 36-38 or 4:428-429.

consonant with recognition of the fact that their exercise can sustain the final value of aims; let us say anything entitled to this sort of treatment *merits respect*.

Claim (3) is used to establish that the ability to sustain the final value of aims is itself a distinctive mode of value. If this is right, then we must respond appropriately to this mode of value. It is a further and more controversial claim to assert that the rational response to this mode of value is Kantian respect, which consists in such things as stable dispositions to refrain from coercing, deceiving or destroying the object of respect and in a stable disposition to treat the happiness of the object of respect as valuable. It would take us too far afield to investigate these claims about the substance of morality, although a complete defense of the formula of humanity would have to establish them. The claim defended here is that there is a substantive requirement, expressed by the formula of humanity, to respect anything that can help sustain final value.³¹

The reconstructed argument proceeds with the observation that there is nothing special about the exercise of one person's rational capacities that should make them differ in value from those of others. It infers from this that all persons merit respect:

(4) If a complete explanation of the rationality of a person's pursuing some aims entails that she merits respect by virtue of her possession of rational capacities, then a complete explanation of the rationality of her pursuing these aims entails that any individual with rational capacities merits respect.

This stage of the realist interpretation of the reconstructed argument differs crucially from the regress argument. Much criticism of the regress focuses on Korsgaard's attempt to move from the claim that one's own humanity merits respect to the claim that the humanity of others also merits respect; she has been charged with begging the question against an egoist who entertains skeptical hypotheses that would vindicate the view that only *his* rational capacities are a source of value. This makes sense, in view of her ambitions. Korsgaard purports to establish not only that her argument succeeds, but also that it is an application of the principle of sufficient reason which can rebut skeptical challenges. That is a tall order, and Korsgaard's commentators have appropriately questioned whether she pulls it off.³²

³¹ One could object that even if the realist interpretation of the reconstructed argument is able to establish the irrationality of violating the formula of humanity, it fails to establish an *obligation* not to violate the formula, and so fails to establish an entitlement against violations of the formula. This objection might maintain that the argument runs afoul of Stephen Darwall's dictum "second-personal authority out, second-personal authority in" or that it illicitly infers claims of what Michael Thompson calls "dikaiological" form from claims of what he calls "monadic" form. See Darwall (2006), 57-60, and Thompson (2004). It would take us too far afield to investigate these issues here. This essay attempts to show that there is a categorical rational requirement not to violate the formula of humanity; a full defense of the claim that this requirement is best understood as an obligation must be taken up elsewhere.

³² See, in particular, Gibbard (1999) and Ridge (2005).

This stage can also be recast, however, as an inference to the best explanation; and once we do this, the realist interpretation of the reconstructed argument avoids these objections to the regress argument. In fact, once it is recast as an inference to the best explanation, this stage of the argument is wholly innocuous. It would be patently absurd for an egoist to claim that his rational capacities can sustain the final value of his aims but to deny that your rational capacities can sustain the final value of your aims, if he supports the former claim with an inference to the best explanation of the final value of his aims. This would be a mistake in explanatory reasoning akin to postulating without motivation that, while all observed electrons have a negative charge that explains their attraction to protons, there are other electrons which lack this charge and so are not attracted to protons. What, after all, is the more plausible explanatory hypothesis: that rational capacities in general are able to sustain the final value of actual choiceworthy aims, and that, as a consequence, the egoist's rational capacities can do this – or that the egoist's capacities alone have can sustain value in this way, and that consequently his aims are finally valuable even though others' are not? An egoist who resists the first hypothesis is reduced to the posture of a solipsistic skeptic who stubbornly resists the hypothesis that the behavior of other persons is explained by the fact that they have minds.

If we could treat as a relatively fixed point of inquiry that the aims of others *lack* value, then this inference would be far from obvious. It would then be more reasonable for the egoist to posit that there is something special about his rational capacities. His position would be like that of a physicist who observes electrons that, for unexplained reasons, fail to be attracted to protons; given this data set, it might well be reasonable to attribute negative charge to some electrons but not to others. But it is not reasonable to begin with an unmotivated skepticism about the value of others' aims.

From here it is but a small step to the remaining claims of the argument:

(5) If a complete explanation of the rationality of a person's pursuing some aims entails that any individual with rational capacities merits respect, then all persons are obligated not to violate the formula of humanity.

This claim is a gloss of what the formula of humanity consists in. As I indicated above, I will not attempt to argue that the details of Kant's understanding of the formula of humanity are supported by my reconstructed argument; but I do believe something closely akin to this is the case. From claims (1) through (5) it follows that:

(6) A complete explanation of the rationality of a person's pursuing some aims entails that all persons are obligated not to violate the formula of humanity.

Freed from the ambitions of Korsgaard's regress argument – vindicating her constructivism about value and thereby rebutting skeptical challenges – a value realist can mimic her justification of the formula of humanity.

4. The Criterion of Moral Standing

I think that the realist interpretation of the reconstructed argument establishes that we must not violate the formula of humanity.³³ If the argument succeeds, it shows there is something special about rational capacities, which motivates taking logocentrism seriously. The exercise of rational capacities can sustain the final value of a person's choiceworthy aims, thereby specifying major constituents of her well-being. But despite endorsing the distinctive value of rationality, I dissent from logocentrism. My reasons for dissent are, moreover, analogous to those I invoke to vindicate the distinctive value of rationality. That is: I believe an argument similar to the realist interpretation of the argument for the formula of humanity establishes a distinct moral principle, which we might call the "formula of consciousness":

Act so that you treat consciousness, whether in your own animal or in that of any other, never merely as a means but always as an end in itself.

If this principle is valid, then consciousness is a locus of moral standing that is not dependent on anything else for its value. To assert this principle is thus to deny logocentrism, since it entails that consciousness does not depend for its value on the value of rationality.

Below I investigate the content of the formula of consciousness in greater detail. But first I present my argument for it:

(1*) Every person has reasons to promote her well-being.

(2*) If a person has reasons to promote her well-being, her capacity for consciousness helps generate these reasons, since a complete explanation of why she has these reasons must invoke this explanatory hypothesis.

(3*) If a complete explanation of why a person has reasons to promote her well-being must invoke the explanatory hypothesis that her capacity for consciousness helps generate these reasons, then a complete explanation of why a person has reasons to promote her own well-being entails that she is entitled to be treated only in ways that are consonant with recognition of the fact that her capacity for consciousness helps generate these reasons; let us say that anything entitled to this sort of treatment *merits concern*.

(4*) If a complete explanation of why a person has reasons to promote her well-being entails that she merits concern by virtue of her capacity for consciousness, then a complete explanation of why a person has reasons to promote her well-being entails that any individual with the capacity of consciousness merits concern.

(5*) If a complete explanation of why a person has reasons to promote her well-being entails that any individual with the capacity for consciousness merits concern, then every person is obligated not to violate the formula of consciousness.

³³ For a different strategy for defending Kantianism as compatible with realism, see David Sussman (2003).

(6*) Every person is obligated not to violate the formula of consciousness.

This argument seeks to capture Wood's focus on animals' own capacities as the source of their moral standing, but to dispense with his claim that these capacities are sources of value only because of their relationship to rationality. The argument seeks to capture Korsgaard's claim that animal capacities, including chiefly consciousness, are a locus of moral standing, but to dispense with her claim that this standing is conferred by the exercise of rationality. It seeks to preserve, in the case of obligations toward animals, the close connection between acting wrongly and making an exception of oneself that is a hallmark of Kantianism. But it seeks to deny logocentrism and to maintain instead that there are at least two loci of moral standing.

I will not investigate in detail the implications of the formula of consciousness, but I would note a few of its features. Firstly, what the formula of consciousness rules out in the first instance is *indifference* to the well-being of conscious individuals. Secondly, the formula of consciousness does not directly rule out killing animals for food or causing severe distress to animals to further scientific research. I take no stance here on whether, when properly interpreted, it prohibits these practices. But it might, for thirdly, it attributes a moral *status* to all conscious individuals, and thus does not support a summative conception of what it is to show concern for conscious individuals. The rationale for the formula of consciousness self-consciously mimics the rationale for the formula of humanity. The case in support of the formula of consciousness thus has, from the outset, a nonconsequentialist structure. It rules out violation of individuals with a certain moral status, and it enjoins the development of a character that is sensitive to possession of that status. Notwithstanding its incorporation of the utilitarian's criterion of moral standing, the formula of consciousness supports neither consequentialism nor optimization of aggregate well-being as a conception of what it is to respond appropriately to the value of well-being.

I now turn to a more specific exposition of the argument. Claim (1*) is not intended to be controversial, so I pass over it without further development.³⁴ I focus on claim (2*):

(2*) If a person has reasons to promote her well-being, her capacity for consciousness helps generate these reasons, since a complete explanation of why she has these reasons must invoke this explanatory hypothesis.

Although this claim looks similar to claim (2) in the realist interpretation of the argument for the formula of humanity, the argument for (2*) is quite dif-

³⁴ A variety of views deny claim (1*), including some forms of nihilism and skepticism. I do not attempt to engage with these views in this essay. I intend (1*) to be compatible with views that deflate well-being by interpreting it as "inclusive"; see, for example, chapter three of Scanlon (1998).

ferent, for the role of consciousness in generating reasons for action is different from the role of rationality. Rationality is claimed to have special status because this hypothesis explains the existence of a certain class of valuable things – namely, finally valuable projects and relationships. By contrast, consciousness is claimed to have special status because this hypothesis vindicates the view that there is reason to promote the well-being of some individuals but not to promote the well-being of other individuals.

To begin my argument for claim (2*), I would echo contemporary Aristotelians who observe that we most commonly speak of conditions being good or bad for plants and animals.³⁵ As these Aristotelians also often observe, furthermore, we commonly speak of conditions being good or bad for artifacts,³⁶ and indeed more generally for anything that is functionally organized, since conditions may conduce to or inhibit its functioning. It is absurd to attribute moral standing to artifacts, however, simply by virtue of the fact that we speak of what is good or bad for them. Similarly, it is absurd to attribute standing to meteorological or geological entities; warm currents are good for hurricanes, and carbon-trapping is good for glaciers, but this does not entail that these things merit our concern.

We thus need a criterion for when the fact that conditions can be good or bad for a thing entails that it is worthy of concern.³⁷ In other words, we need a criterion of moral standing, and the formula of consciousness suggests consciousness as the relevant criterion. In my view the most plausible rival to consciousness as the criterion of moral standing is not the Kantian criterion of rationality, but rather the criterion typically proposed by Aristotelians: life. Richard Kraut advocates this view, for example, and suggests it is evidenced by our application of the term “flourishing” to all and only living things.³⁸ Kraut writes:

To see this, consider a child who plans on lighting a fire and destroying a forest, simply for the sake of such destruction. Should we stop him? Certainly. But why so? – is it only because he may endanger human lives, kill the animals who live in the forest, and prevent it from serving human purposes? Why should not the fact that his act is bad for trees and a great many other living things – in fact, *all* the things in the forest that are flourishing – also be counted as a reason for interfering with him? ... [T]he child is not innocently rearranging the world; he is deliberately inflicting harm on all those forms of life for its own sake. That state of mind should disturb us, and so there must be something objectionable in what he is trying to do. That we do not normally speak

³⁵ See Kraut (2007), including especially sections 2, 12 and 56; and, for an earlier view, see Philippa Foot (2001). Kraut and Foot are both inspired by the classic discussion in Aristotle (c. 350 B.C.E.), including especially Book I.

³⁶ Kraut (2007), 3.

³⁷ Compare Darwall: “What gives someone’s welfare or personal good the status of normative reasons is his having a value that makes him worthy of care.” See Darwall (2002), 8. Earlier allied views include that of Elizabeth Anderson (1993), 8-11, and J. David Velleman (1999).

³⁸ See Kraut (2007), 5-8. I am not sure this claim about ordinary language is correct; it seems to me that we can speak ordinarily of the flourishing of hurricanes and glaciers.

of the welfare of trees should not diminish our concern about what the child does. We should say that since his act is bad for a great many beings (including those that are not classified as members of the animal kingdom), and good for none, that settles the matter: he should not light the fire.³⁹

There certainly is something objectionable about the state of mind of this child. But in my view we cannot infer from this that trees have moral standing. One reason why not is that the child's destructiveness may evidence disregard for conscious individuals even if none of the individuals he destroys are conscious: we should also be concerned about the state of mind of a child who wantonly destroys an abandoned beaver dam, beautiful rock formation, photograph or doll.

The view that plants have moral standing, even in the weak sense that we have reason (for their sake) not to wantonly destroy them, may be thought to entail absurd conclusions, such as that we are not entitled to raise crops for food or use timber for construction. It may appear to entail that Kant's ghastly inference – concluding that animals are mere things to be used at will from the premise that they lack rationality – is equally ghastly when applied to plants. But as Kraut observes, this is not the case. On Kraut's view, in fact, the reasons that are generated by plant moral standing are exceedingly weak. He writes:

Plants should not be wantonly destroyed, but there is no reason to make a special effort simply for the sake of their good. We should benefit them only if that in turn helps someone else who should be helped; they fall below the threshold of merited direct concern. ... all that plants can achieve ... is not enough to make it worthwhile to take action merely to make them better off. There are much better things for us to do.⁴⁰

Kraut's view appears to be that we have reasons to promote the good of plants, but that these reasons are so vanishingly weak that they are always defeated by the opportunity cost of using time and resources to respond to them. This is not a consequence, moreover, of any accident of our circumstances; I would always do better to amuse myself, or to cultivate a plant for my own sake, than to cultivate a plant for its sake. On this view we can therefore, without rational error, deploy a heuristic of ignoring the flourishing of plants. The only time the relevance of plant flourishing is exposed is when, like the child lighting the forest on fire, a person destroys the plant without achieving any good whatsoever.

There are considerations that should give us pause, however, about accepting life as the criterion of moral standing. There is some tension, I believe, in the passages from Kraut that I have reproduced here; and this tension points us toward a difficulty for the view that life is the criterion of

³⁹ Kraut (2007), 7-8. Later he adds that "... if pleasure is not the only good and certain things are good for plants, [a conscientious person] must make plants, no less than conscious animals and human beings, the object of his concern." See Kraut (2007), 42.

⁴⁰ Kraut (2007), 211-212.

standing. The tension in question is between the view that an individual could have moral standing in the sense that its well-being generates reasons not to destroy it and the view that its well-being fails to generate significant reasons to promote its good.

To see what I have in mind, consider the conclusion Kraut reaches in the last quotation: we may as well not have reasons to promote the good of plants, because the opportunity cost of doing so always defeats the weak force of these reasons. Now let us apply this insight back into the case of the child who starts a fire in an uninhabited and non-useful forest for the sake of destroying the trees. On Kraut's view, that the child destroys living (though non-conscious) things figures importantly in explaining why his behavior is objectionable. One could wonder why this is so, however, in light of his claim that plant flourishing generates only exceedingly weak reasons. If we stipulate that the child does no good in burning the forest, then even an exceedingly weak reason may suffice to show that he should not do it. But the last quotation entails there is a more important reason not to burn the forest: the opportunity cost of time and energy. As Kraut says, *there are much better things for us to do*. If this opportunity cost always defeats the flourishing of plants in ordinary deliberation about what to do, then it is a reason not to burn the forest that is much stronger than any reason not to destroy the plants as such. This putative latter reason, whether or not it exists, thus fails to inform the child's deliberation about whether to burn the forest. That we concur with the judgment that the child should not burn the forest thus fails to establish that the plants in the forest have moral standing, for there is a more powerful reason – the opportunity cost of pointless activity – that vindicates this conclusion. It is difficult to sustain the significance of marking a distinction between meriting *direct* concern and meriting concern more generally.

One might reply that there are stronger reasons not to destroy living things than there are to promote their flourishing. But in advancing this reply there is danger of recovering the absurd putative entailments of attributing plant moral standing – that, in at least some circumstances, it is impermissible *for their sake* to destroy them for food or shelter – that Kraut avoids by claiming that our reasons to promote plant flourishing are exceedingly weak.

Or one might reply that destruction is an especially important category because it is evidence of bad character. Eliminating or mastering destructive inclinations not only prevents people from doing bad things, it also constitutes moral education and self-improvement. But this route to the significance of destruction deemphasizes the plants themselves: it is the character of the child who destroys plants, not the fact that living things are destroyed, that rightly dominates our concern. Thus, even if this observation is correct, it fails to motivate plant moral standing.

The advocate of life as the criterion of moral standing could concede the force of this argument yet still maintain that the fact that trees are living things is a distinct cause for concern which is absent when the child destroys

stuffed animals. This strategy concedes, however, that the good of the plants themselves never enters significantly into practical deliberation. Such an attribution of moral standing is so vanishingly weak that it is not clear if it makes sense to call it standing, or to say that responding to it involves a show of concern. It is extremely difficult to adjudicate between that position and one where the criterion of moral standing is consciousness; when put into practice, the former position collapses into the latter. Thus, despite the plausibility in the hypothesis that being alive is the criterion of moral standing, I believe consciousness is the relevant criterion, and consequently I endorse claim (2*) above.⁴¹

The last three steps of my argument for the formula of consciousness should be relatively uncontroversial. Claim (4*) is a consequence of the same inference that vindicates claim (2*): an individual's consciousness makes its well-being reason-generating because this hypothesis best explains why my well-being is reason-generating though the well-being of pianos and hurricanes is not. This parallels the relationship between claims (2) and (4) in my realist argument for the formula of humanity. Claim (5*) follows from the previous four claims and the definition of the formula of consciousness, and claim (6*) is entailed by the previous five claims.

The chief obstacle to the success of the argument, apart from whether it succeeds in vindicating claim (2*), is whether it is correct in claim (3*) to characterize the reason-generating status of consciousness as an *entitlement* of conscious individuals, where this entails that failing to respond appropriately to this status involves acting wrongly. I am not able here to provide a conclusive defense of this claim, but it is worth noting again here the similarities between the rationale for the formula of consciousness and the constructivist's rationale for the formula of humanity. The rationale for the formula of consciousness is as follows: since the rationality of every action depends on attributing a moral status to individuals possessing the capacity of consciousness, to be rational, persons must regulate their actions by responding to that moral status. Whenever a person violates the formula of consciousness, that is, he fails to attribute a normative status to one conscious individual that he must attribute to another to explain fully the rationality of his actions. A complete explanation of the rationality of any action, that is, must appeal to at least one individual's consciousness as generating reasons. Failing to value another individual's consciousness in action – failing to have *concern* for that individual, in the technical sense of that term I deploy here – thus treats at least one conscious individual's capacities as exceptional, and hence elevates that individual with respect to another with the same capacities. This exception-making, this differential treatment of individuals with the same moral

⁴¹ Here I follow Anderson: "... there is a base line of care which we should show for all animals. I suggest that we call this kind of valuation 'consideration.' Consideration is a way of caring which pays due regard for the interests of conscious beings, apart from whether they are rational." See Anderson (1993), 9-10.

status, is the mark of wrongful action in the case of a failure to respect an individual's status as a rational individual. Such a mistake is not a mere failure to respond appropriately to reasons or values, but is more strongly a violation or denigration of her; that is why it is appropriate to characterize her claims against such treatment as entitlements. The formula of consciousness advances a parallel claim in the case of the status individuals have in virtue of their consciousness. This parallelism supports the claim that the formula of consciousness captures not only an important class of reasons but also an important class of *entitlements* of individuals, and hence also an important class of obligations to heed these entitlements.

5. The Formula of Consciousness

If my argument succeeds in its aspirations, then all persons are obligated not to violate the formula of consciousness:

Act so that you treat consciousness, whether in your own animal or in that of any other, never merely as a means but always as an end in itself.

I label this principle the “formula of consciousness” to emphasize its similarity to the formula of humanity. This principle contains a qualification akin to the personification principle, which we might call the “animalification principle.” This clause limits the requirement to show concern for conscious individuals to cases where consciousness is found in animals. This qualification in the formula of consciousness prevents the use of the principle, by analogy to Wood's extension of Kant's formula of humanity, to justify obligations in regard to individuals with “fragments” or the “infrastructure” of consciousness. Mere possession of a central nervous system or of the capacity to move in response to one's environment, for example, fails to entitle an individual to concern: only full-blown consciousness suffices for moral standing.

I follow Wood in dropping the personification principle from the formula of humanity. But unlike Wood, as the argument above exposes, I suggest we replace it with a new criterion of moral standing. Contrary to both Kant's logocentric endorsement of the personification principle and Wood's logocentric rejection of this principle, I suggest consciousness as the relevant criterion. In expressing the formula of humanity, then, I recommend replacing the personification principle with the animalification principle:

Act so that you treat humanity, whether in your own *animal* or in that of any other, never merely as a means but always as an end in itself.

If we stipulate that we are using the term “humanity” as Kant uses it, the animalification principle is redundant in this formulation. Humanity is rationality as instantiated in embodied beings, and all embodied rational beings are conscious animals. It is nevertheless important to articulate the animalification principle explicitly, for it calls attention to the fact that humanity is com-

plex, encompassing both our rationality and our consciousness. In my view the notion of humanity must be analyzed into these respective components if we are to provide the most perspicuous account of the obligations we owe to rational individuals. I thus recommend that the formula of humanity be amended as follows:

Act so that you treat *rationality*, whether in your own *animal* or in that of any other, never merely as a means but always as an end in itself.

In this formulation, the animalification principle is *not* redundant. It distinguishes cases where rational capacities are instantiated in conscious individuals (such as human persons) from cases, should there be any, where they are instantiated in non-conscious individuals (collective agents, perhaps). The first clause tells us what we must treat as valuable (rationality), and the second tells us when we must so treat it (when found in conscious individuals). This is compatible with *also* pursuing Wood's strategy of attributing moral status to animals under the formula of humanity; I take no position on whether such an attribution is warranted.

To more clearly distinguish the formula of humanity from the formula of consciousness, I recommend further that the rationale for each formula be incorporated into its expression, along with a brief gloss of how it may be satisfied. Thus the formula of humanity reads:

In recognition of the fact that the exercise of rational capacities can sustain the final value of an individual's choiceworthy aims, act so that you treat rationality, whether in your own animal or in that of any other, never merely as a means but always as an end in itself, by demonstrating respect for every individual who possesses this capacity.

And the parallel formula of consciousness reads:

In recognition of the fact that the exercise of the capacities of consciousness generates reasons to promote an individual's well-being, act so that you treat consciousness, whether in your own animal or in that of any other, never merely as a means but always as an end in itself, by demonstrating concern for every individual who possesses this capacity.

Each of these principles encompasses an important class of moral obligations. In the concluding section, I briefly explore some consequences of accepting both principles.

6. Conclusion

In this final section I note two implications of the account of morality I have sketched and argued for in this essay. I begin with the claim that consciousness is a locus of moral standing in its own right, not dependent on its relationship to rationality for its ability to generate reasons and obligations. Since

non-rational animals lack the capacity to cognize reasons as reasons, however, they are not themselves subject to moral norms. I thus depart from Kant's view and other logocentric views, Wood's and Korsgaard's included, by denying the following claim:

Moral Community Closure: An individual is subject to moral norms just in case she is herself an ultimate source of moral norms.⁴²

The rider "ultimate" is crucial here, since both Wood and Korsgaard contend that there is a sense in which animals are a source of moral norms. But their logocentrism commits them to the view that animal moral standing is derived from the moral standing of rational individuals.

Although the view I defend is not logocentric, it privileges rationality in a variety of respects.⁴³ Rationality is held to be a distinct locus of moral status, when found in conscious individuals; rational conscious individuals merit respect, non-rational conscious individuals do not. This entails that rational conscious individuals are of a higher moral standing than their non-rational counterparts. This is important, since one of the chief virtues of the position is that it preserves Kantian insights about the distinctive moral status of persons. It is also independently plausible, since it explains why conscious persons should be accorded priority over non-rational animals in circumstances of triage. I thus deny:

Moral Standing Egalitarianism: If x and y are individuals with moral standing, then x 's moral standing is equal to y 's moral standing.

It may seem obvious that this claim is false, since both persons and animals have moral standing, yet persons have a higher moral standing than animals. Practitioners of both Kantianism and utilitarianism often write, however, as though they believe this claim. Kantians sometimes deny that animals have moral standing, and utilitarians often assert that the priority appropriately given to persons is a consequence only of the kinds of goods they are able to enjoy.⁴⁴ In this essay I have begun to show how an important insight of each of these leading traditions of modern moral theory – the Kantian insight that rational conscious individuals have a higher status than non-rational conscious individuals, and the utilitarian insight that consciousness is the criterion of moral standing⁴⁵ – can be incorporated into a unified theory. I have

⁴² I thus also deny *Moral Community Symmetry*: If x is subject to moral norms as a result of y 's moral standing, then y is subject to moral norms as a result of x 's standing.

⁴³ This feature is what makes this approach more appealing, in my view, than the influential approach of Tom Regan; but I do not have space to consider Regan's view here. See Regan (1983).

⁴⁴ The classic statement of this view is found in chapter two of John Stuart Mill (1861).

⁴⁵ This claim appears throughout the utilitarian tradition; classic statements of it include Jeremy Bentham (1781), especially chapter one, Mill (1861), especially chapter two, Henry Sidgwick (1874), especially chapter one of Book IV, and Singer (1974).

not fleshed out this theory in all its details, of course, and in particular I have not provided an account of how these principles are to be simultaneously satisfied. I have attempted to articulate a plausible framework, however, within which this important project may be pursued. If this attempt to articulate a new framework is successful, that is significant, for it promises an avenue through which to move beyond many of the disputes between Kantians and utilitarians that have characterized, and also to an extent ossified, modern moral theory.

Jon Garthoff
Northwestern University
Department of Philosophy
garthoff@northwestern.edu

References

- Ameriks, Karl (1989). "Kant on the Good Will." In *Grundlegung zur Metaphysik der Sitten*. Otfried Höffe (ed.). Klostermann.
- Anderson, Elizabeth (1993). *Value in Ethics and Economics*. Harvard.
- Aristotle (c. 350 B.C.E.). *Nicomachean Ethics*. Terence Irwin (trans. and ed.). Hackett, 1999.
- Bentham, Jeremy (1781). *The Principles of Morals and Legislation*. Prometheus, 1988.
- Bratman, Michael (2007). *Structures of Agency*. Oxford.
- Cavalieri, Paola and Peter Singer (1993). *The Great Ape Project*. St. Martin's.
- Darwall, Stephen (2002). *Welfare and Rational Care*. Princeton.
- _____ (2006). *The Second-Person Standpoint*. Harvard.
- FitzPatrick, William J. (2005). "The Practical Turn in Ethical Theory." *Ethics* 115.
- Foot, Philippa (2001). *Natural Goodness*. Oxford.
- Frankfurt, Harry (2004). *The Reasons of Love*. Princeton.
- Garthoff, Jon (2010a). "Structuring Ends." *Philosophia* 38:4.
- _____ (2010b). "Mimicking Korsgaard." In *Rethinking Kant: Volume Two*. Pablo Muchnik (ed.). Cambridge Scholars.
- Gaut, Berys (1997). "The Structure of Practical Reason." In *Ethics and Practical Reason*, Garrett Cullity and Berys Gaut (eds.). Oxford.
- Gibbard, Allan (1999). "Morality as Consistency in Living." *Ethics* 110.
- Hare, R. M. (1963). *Freedom and Reason*. Oxford.
- Kant, Immanuel (1785). *Groundwork of the Metaphysics of Morals*. Mary Gregor (trans. and ed.). Cambridge, 1997.
- _____ (1786). *Conjectures on the Beginning of Human History*. In *Kant: Political Writings*. Trans. H. B. Nisbet. Hans Reiss (ed.). Cambridge, 1970.
- _____ (1793). *Religion Within the Boundaries of Mere Reason*. In *Religion and Rational Theology*. Allen W. Wood and George Di Giovanni (trans. and eds.). Cambridge, 1996.
- _____ (1797). *The Metaphysics of Morals*. Mary Gregor (trans. and ed.). Cambridge, 1996.
- _____ (1798). *Anthropology from a Pragmatic Point of View*. Robert B. Louden (trans. and ed.). Cambridge, 2006.
- Korsgaard, Christine (1983). "Two Distinctions in Goodness." In *Creating the Kingdom of Ends*. Cambridge, 1996.
- _____ (1986a). "Kant's Formula of Humanity." In *Creating the Kingdom of Ends*. Cambridge, 1996.
- _____ (1986b). "Aristotle and Kant on the Source of Value." In *Creating the Kingdom of Ends*. Cambridge, 1996.
- _____ (1996a). *The Sources of Normativity*. Onora O'Neill (ed.) Cambridge.
- _____ (1996b). "An Introduction to the Ethical, Political, and Religious Thought of Kant." In *Creating the Kingdom of Ends*. Cambridge.
- _____ (2003). "Realism and Constructivism in Twentieth-Century Moral Philosophy." *Journal of Philosophical Research*.
- _____ (2004). "Fellow Creatures." In *The Tanner Lectures on Human Values* Vol. 25/26. Grethe B. Peterson (ed.). Utah.
- _____ (2007). "Facing the Animal You See in the Mirror." Panel presentation delivered at Harvard University on April 24, 2007. Forthcoming in *The Harvard Review of Philosophy*.
- _____ (2009). *Self-Constitution*. Oxford.
- Kraut, Richard (2007). *What Is Good and Why*. Harvard.
- Langton, Rae (2007). "Objective and Unconditioned Value." *Philosophical Review* 116.
- Mill, John Stuart (1861). *Utilitarianism*. Roger Crisp (ed.). Oxford, 1998.
- O'Neill, Onora (1998). "Kant on Duties Regarding Nonrational Nature II." *Proceedings of the Aristotelian Society* Suppl. Vol. 72.
- Regan, Tom (1983). *The Case for Animal Rights*. University of California, 1985.

- Richardson, Henry (1997). *Practical Reasoning About Final Ends*. Cambridge.
- Ridge, Michael (2005). "Why Must We Treat Humanity With Respect?." *European Journal of Analytic Philosophy*.
- Scanlon, T. M. (1998). *What We Owe to Each Other*. Harvard.
- Schmidtz, David (1994). "Choosing Ends." *Ethics* 104.
- Sidgwick, Henry (1874). *The Methods of Ethics*. Hackett, 1981.
- Singer, Peter (1974). "All Animals Are Equal." *Philosophic Exchange* 1.
- Sussman, David (2003). "The Authority of Humanity." *Ethics* 113.
- Thompson, Michael (2004). "What Is It to Wrong Someone?" In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*. Jay Wallace, Philip Pettit, Samuel Scheffler and Michael Smith (eds.). Oxford.
- Velleman, J. David (1999). "A Right of Self-Termination?" *Ethics* 109.
- Wood, Allen W. (1998). "Kant on Duties Regarding Nonrational Nature I." *Proceedings of the Aristotelian Society* Suppl. Vol. 72.
- _____ (1999). *Kant's Ethical Thought*. Cambridge.
- _____ (2008). *Kantian Ethics*. Cambridge.