*The Journal of Value Inquiry* **36**: 349–367, 2002.



# O'Neill and Korsgaard on the Construction of Normativity

ERIC WATKINS[1] AND WILLIAM FITZPATRICK[2]
[1]*Department of Philosophy, University of California, San Diego, La Jolla, CA 92093-0119, USA;* [2]*Department of Philosophy, Virginia Tech, Blacksburg, VA 24061-0126, USA*

Over the past two decades it has become standard to construe Kant's ethics as a paradigm of constructivism, standing in stark opposition to realism. Under such an interpretation, Kant's ethics is based on constructive procedures that yield normative practical principles for us, rather than on facts about goodness that might ground such principles. While there is, no doubt, textual support for taking Kant to be specially concerned with such procedures, the motive for construing his ethical theory as constructivist is largely philosophical. Adopting such a constructivist interpretation would allow us to avoid the metaphysical commitments of realism, thus side-stepping the need to provide a metaphysical defense of the values in question, not to mention the difficulties of trying to explain how objective, non-constructed values fit into the metaphysics of the modern world and are consistent with the modern scientific worldview.[1] Onora O'Neill and Christine Korsgaard have developed detailed constructivist interpretations of Kantian ethics, which they believe are more satisfactory than realist alternatives. However, both versions of constructivism encounter significant philosophical difficulties, making it worthwhile to consider what a more realist approach to Kant's ethics might look like.[2]

## 1. O'Neill's Critique of Rawls's Constructivism

Onora O'Neill develops her constructivist interpretation and reconstruction of Kant's ethics in *Constructions of Reason: Explorations of Kant's Practical Philosophy* and *Towards Justice and Virtue: A Constructive Account of Practical Reasoning*. In order to see what is distinctive about O'Neill's version of constructivism, it will be helpful to see how it contrasts with the view of John Rawls, since it was Rawls who first made the term "constructivism" popular in political philosophy in *A Theory of Justice*. In "Constructivisms in Ethics," chapter eleven of *Constructions of Reason*, O'Neill presents Rawls's constructivism as an attempt to chart a middle path between realism and relativism. Rawls's constructivism is anti-realist insofar as Rawls attempts to steer clear of any transcendent metaphysical claims, but, at the same time, it is

supposed to avoid relativism by determining principles of justice that everyone can accept. As Rawls explains the basic idea of his theory, "it specifies a particular conception of the person as an element in a reasonable procedure of construction, the outcome of which determines the content of the first principles of justice," and it is intended "to establish a suitable connection between a particular conception of the person and first principles of justice, by means of a procedure of construction."[3] Rawls is careful to distinguish his statement of Kantian constructivism from Kant's own views: "Justice as fairness is not, plainly, Kant's view, strictly speaking; it departs from his text at many points."[4] Rawls calls the means of establishing a connection between a particular conception of the person and first principles of justice "constructive," because the connection is not supposed to be established by means of a

> search for moral truth interpreted as fixed by a prior and independent order of objects and relations, whether natural or divine, an order apart and distinct from how we conceive of ourselves. . . . What justifies a conception of justice is not its being true to an order antecedent to and given to us, but its congruence with our deeper understanding of ourselves. . . . Kantian constructivism holds that moral objectivity is to be understood in terms of a suitably constructed social point of view that all can accept.[5]

In filling out his constructivist account, Rawls develops a powerful set of tools to establish the connection between a conception of the person and the first principles of justice. In particular, he thinks that such principles can be established by means of a process of reflective equilibrium from within the original position, which is defined by the veil of ignorance it places on agents. The basic idea is that first principles of justice can be constructed by determining what agents would choose upon reflection from behind the veil of ignorance, their situation if they knew neither who they would be nor what their particular desires would be. Since the appeal to the veil of ignorance rules out any specific information that such agents might use to gain unfair advantage for themselves, whatever principles they would choose in such a situation would have to be fair and thus just. However, the veil of ignorance cannot exclude all information and desires, since in that case agents in the original position would have no guidance in selecting any principles at all, much less fair or just ones. As a result, Rawls breaches the veil of ignorance to allow knowledge of the universal desirability of primary goods such as rights, liberties, and wealth into the original position. Given this carefully devised knowledge base, agents in the original position can reflectively construct the first principles of justice.

O'Neill begins her critique of Rawls by asking what justification he has for the conception of the person as defined in the original position. Someone might simply reply that this conception is part of, or implied by, an adequate

philosophical account of human nature. Rawls, however, avoids such a realist answer because he wants to eschew, as O'Neill puts it, "an account, let alone the suspect Kantian account, of the metaphysics of the self," or, more pejoratively, "obscure and panicky metaphysics."[6] Accordingly, Rawls ultimately favors the alternative of admitting that "we are not trying to find a conception of justice suitable for all societies regardless of their particular social or historical circumstances."[7] However, as O'Neill critically remarks, "far from deriving a justification of democratic citizenship from metaphysical foundations, Rawls [can vindicate only] those deep principles of justice 'we' would discover in drawing on 'our' underlying conceptions of free and equal citizenship. This vindication of justice does not address others who, unlike 'us', do not start with such ideals of citizenship; it has nothing to say to those others. It is 'our' ideal, and 'our' justice."[8] In short, according to O'Neill, Rawls's constructivism is not ultimately distinct from the relativism it was designed to avoid.[9]

Where, according to O'Neill, does Rawls go astray? Given that O'Neill, too, is attracted to constructivism, it is crucial that she identify and then eliminate the feature of Rawls's particular version of constructivism that prevents his account from being satisfactory. According to O'Neill, Rawls's difficulty stems from the fact that his version of construction involves an ideal of the person in the original position as opposed to a mere abstraction from certain facts about persons. As O'Neill puts it: "Idealization masquerading as abstraction produces theories that may appear to apply widely, but in fact covertly exclude from their scope those who do not match a certain ideal. They privilege certain sorts of human agent and life by presenting their specific characteristics as universal ideals."[10] Accordingly, once Rawls has idealized agents in the original position, the only way in which he can justify its implications is by restricting its application to western democratic societies.[11]

## 2. O'Neill's Positive Constructivist Account

If O'Neill is correct in thinking of Rawls's constructivism as an idealizing constructivism insofar as he invokes an ideal of the person in the original position that may not apply to all, what does she think the fundamental features of a more adequate constructivism, a non-idealizing constructivism, ought to be? Instead of invoking a veil of ignorance that has been selectively breached so as to let into the original position a limited set of desires whose satisfaction is to be maximized by instrumental reason, O'Neill abstracts from desires altogether, suggesting that the operative principle of construction stems from answering the modal question: "What principles can a plurality of agents of minimal rationality and indeterminate capacities for mutual independence live by?"[12] O'Neill contrasts this question with the hypothetical question that

would reflect a more Rawlsian approach, "What principles would a plurality of [such] agents choose to live by?", insofar as answering the Rawlsian question depends on choosing which desires to grant agents, thereby generating the question's hypothetical character.

If desires are not the crucial determinant within O'Neill's constructivist procedure, what is? And if instrumental rationality depends on desires and desires have been deemed irrelevant to the constructive procedure, what conception of rationality is at work in constructivism? The fundamental notion for O'Neill is the possible agency of a plurality of distinct, but interrelated rational beings. O'Neill's account can generate normative practical principles because "there are certain constraints on the principles of action that could be adopted by all of a plurality of potentially interacting agents of whom we assume only minimal rationality and indeterminate mutual independence. Principles that cannot be acted on by all must be rejected by any plurality for whom the problem of justice arises."[13] To illustrate her idea, O'Neill shows how principles of coercion, violence, and deception can be shown to be violations of the first principles of justice, given that they cannot be acted on or accepted by all, since coercion, violence, and deception undercut the rational agency or assent of at least some agents, namely that of those who would be coerced, harmed, or deceived on such principles. Instead of invoking instrumental reason in order to satisfy as many of a selective set of desires as possible, O'Neill's version of constructivism is constructed to ensure that it is possible for agents to be aware of, consent to, and act on the practical principles that are to be constructed. By beginning with an account of agents that abstracts from their desires, but leaves a minimal account of rationality intact, and by considering what principles it is possible for all such agents to act on, O'Neill ensures that her version of constructivism is distinct from Rawls's in significant ways.

According to this description of O'Neill's account, the problem of relativism that Rawls's account faces can apparently be avoided without lapsing into any unacceptable metaphysical claims. Whereas Rawls was forced to accept relativism insofar as the principles that resulted from his constructivist procedure would be accepted only by those who share the ideal conception of an agent built into its foundation, O'Neill's account would apply to any rational agents that we could possibly interact with. At the same time, she need not appeal either to any controversial metaphysical claims or to an unvindicated ideal of the person. Thus, it might seem as if O'Neill has developed a version of constructivism that can be seen as having significant advantages over Rawls's, without any apparent drawbacks.

However, precisely because O'Neill does not idealize with her version of constructivism in the way that Rawls does, she faces a significant challenge where he does not, namely in explaining the normative force of the principles that result from the constructive procedure she advances. While Rawls can

account for the normative force of practical principles insofar as they follow from Western society's acceptance of the ideals implicit in the original position, O'Neill, who rejects his idealization, cannot. If practical principles have normative force and these principles are the result of constructive procedures, then the constructive procedures must be responsible for generating the normative force. Yet the constructive procedures cannot generate normative force if they do not already possess normative authority. At the same time, it is unclear how a constructivist could claim that constructive procedures have normative authority intrinsically, if they are based simply on certain general modal facts.

O'Neill addresses this challenge in "Reason and Politics in the Kantian Enterprise" and "The Public Use of Reason," the first two chapters of *Constructions of Reason*, as well as in "Practical Reason: Abstraction and Construction," the second chapter of *Towards Justice and Virtue*. She argues that we can see how the constructive procedure can produce normative force for its resultant principles by pursuing the following line of thought. If the constructive procedure is based on reason, then the principles constructed by such a rational procedure can have normative force only if reason has the normative authority required to generate such force. But in considering whether reason has such a normative authority, the following dilemma that faces any justification of the authority of reason naturally arises. Either something other than reason justifies the authority of reason, or reason must vindicate its own authority. If something alien to reason is supposed to justify its authority, then it too would stand in need of vindication and we would have made no progress since we could always ask whether that factor on a particular occasion was rationally justified. However, appealing to reason itself might seem to be clearly circular. How could reason possibly justify its own authority?[14]

O'Neill responds to this dilemma by arguing that the authority of reason can be justified, but only recursively and in a public or political context. O'Neill suggests that we should take seriously both the notion of the public use of reason Kant employs in his essay "What is Enlightenment?" and the political metaphors scattered throughout Kant's *Critique of Pure Reason*. According to O'Neill, the public use of reason is a use of reason that does not appeal to any external authority, any authority that could not be accepted by all of a plurality of rational agents. Thus, regardless of how extensive his audience, a priest's statement from the altar is not public in O'Neill's sense since the priest speaks from the standpoint of someone who accepts an authority other than reason, and to the limited audience that accepts such an authority. Accordingly, the only authority that can be appealed to during the public use of reason is one that is implicit in or internal to the standards of public or political discourse.

What speaks in favor of the public use of reason rather than some other aspect of rationality in constructing normative principles? According to

O'Neill, no other justification is available. A more fully metaphysical justification of the normative force of reason is ruled out on the basis of her rejection of transcendent metaphysics: "There is no lofty position above the debate, as perhaps there might be if human reason had a transcendent source."[15] A theological justification is ruled out given the fact that agents are independent and autonomous:

> no master plan is inscribed in each one of us; rather we must devise a plan. . . . This plan must not presuppose unavailable capacities to coordinate, such as a preestablished harmony between reasoners. . . . The most basic requirement for construction by any plurality of agents must be . . . that any fundamental principles of thought and action we deploy be ones that it is not impossible for all to follow.[16]

In fact, any kind of justification that is not based on the standards inherent in public debate is ruled out: "since the world at large accepts no common external authority, the only authority the communication can assume must be internal to the communication."[17] As a result, "reason, on this account, has no transcendent foundation, but is rather based on agreement of a certain sort. Mere agreement, were it possible, would not have any authority. What makes agreement of a certain sort authoritative is that it is agreement based on principles that meet their own criticism."[18] In short, reason, and thus the practical principles based on it, can be justified on O'Neill's account, but only recursively by means of the standards that can be accepted or agreed upon by a plurality of rational agents engaging in public or political discourse.

## 3. Criticisms of O'Neill's Account of Normativity

What thus seems to motivate O'Neill's account is the following line of thought. Since people disagree about metaphysical claims, presumably because metaphysical entities transcend what human beings can know, and since agreement is crucial in being able to solve the practical problems of morality and justice, O'Neill dismisses metaphysical principles in search of constructive principles that can secure agreement as she attempts to develop an adequate account of the normativity of the practical principles of morality and justice. The particular version of constructivism that O'Neill then offers, a recursive, political account that is based on the public use of reason, can explain, she thinks, how the kind of agreement that is essential to solving the practical problems of morality and justice can be generated without appealing to any metaphysical principles, because the public use of reason is conceived of in terms of public agreement and does not rely on any metaphysical claims that would transcend what human beings can know.

However, can O'Neill's version of constructivism contain a rich enough account of the normativity that we typically ascribe to the practical principles of morality and justice? In order to see what might be missing in O'Neill's account, consider the structure of her argument for the connection between reason and the normative force of practical principles. In order to justify the normative force of rational practical principles, we must appeal to either an external or an internal authority. Since there is no agreement on any external authority such as God or transcendent values, if there is to be any authority, it must be an internal authority. The only internal authority that seems at all plausible is derived from the public use of reason, which demands that moral principles stem from agreement that is "based on principles that meet their own criticism."[19] Since arguments that proceed negatively by eliminating other alternatives cannot be conclusive, if they do not also establish positively that the remaining alternative is sufficient to do what is required of it, even if O'Neill has argued or, from a metaphysical standpoint, perhaps simply presupposed, negatively that external authorities cannot be used to generate normativity, it is still incumbent upon her to show positively that the public use of reason can generate the requisite sense of normativity. O'Neill's positive account invokes a certain feature of reason, the possibility of following its principles by all rational agents, as its defining or essential feature. As she puts it: "'reason' is just the name we give to whatever may be most authoritative for orienting thought and action."[20] In short, O'Neill defines "reason" in terms of universally followable principles and they are what must be used in explaining the authority of reason and thus the normativity of practical principles.

Yet it may not be fully evident that O'Neill's positive account is successful in establishing the sense of normativity that we typically ascribe to the practical principles of morality and justice. First, it is unclear that possible agreement in public discourse is sufficiently rich to capture the full content of normativity. In Section IV of the General Introduction to *The Metaphysics of Morals*, Kant introduces grounds of obligation and argues that only grounds of obligation, not obligations themselves, can be opposed to each other. When grounds conflict, Kant says, the stronger of the two grounds generates an obligation, while the weaker does not bring about any obligation at all. Thus, according to Kant, it seems to be the case both that grounds of obligation can conflict and that obligations come in various degrees since what obligations arise depends on the various strengths of the different grounds involved. Yet both of these views would seem to be problematic for O'Neill's account, given that modal facts about the possibility of following do not seem to admit of conflict or degrees. Similarly, it is not clear that the mere fact that people can follow a certain principle immediately establishes it as a principle of morality or justice in particular. It seems that we all can follow some principles such as *modus ponens*, even principles concerning our actions such as hypotheti-

cal imperatives, that we do not think of as belonging to morality or justice proper. In order to establish rather than merely suggest these points, various senses of normativity would need to be spelled out and, in particular, an account that is stronger than O'Neill's would need to be articulated without begging the question against her. While actually carrying such an explanation out in detail extends beyond the scope of our discussion, it is certainly a question that can be raised against O'Neill's account.

Second, the characterization of reason that O'Neill explicitly appeals to in her positive account departs from our common sense conception of the nature of reason. One natural reaction to O'Neill's account of reason is to say that it would have us inappropriately define reason in terms of one of its consequences rather than in terms of its nature. It might be uncontroversial to assert that if something is rational, then everyone can follow it. However, the same cannot be said of O'Neill's claim that what makes something rational is the fact that everyone can follow it. It is more plausible to say that it is something about the proposition in question that determines whether it is rational or not and that it is merely a consequence of the proposition displaying that feature that everyone can follow it. Moreover, O'Neill's account of reason cannot simply be replaced with a different account of reason, because what is distinctive about O'Neill's version of constructivism stems directly from her account of reason. It is only by equating rationality and the possibility of following that she is warranted in viewing her account as the particular version of constructivism that it is. Since the possibility of following is constructed in a public sense, the authority of reason and of the practical principles based on it must be constructed as well. If such an identification of reason and the possibility of following is rejected, the rationale in favor of understanding it as a version of constructivism disappears.

In light of these points, it is worth noting briefly that Kant's own broader account of rationality seems to provide precisely what must be foreign to O'Neill's view, an account of the nature of reason. It is a defining feature of Kant's account of reason that reason seeks the unconditioned condition of everything conditioned. Moreover, in conjunction with this account, Kant also accepts principles, called Ideas of reason, which he takes to be regulative for our theoretical and practical behavior. On the theoretical side, Kant argues in the Appendix to the Transcendental Dialectic of the *Critique of Pure Reason* that we are supposed to proceed as if God created the world, as if the world were an absolute totality of substances, and as if we were simple souls, where God, the world, and our souls are unconditioned conditions, incapable of being given to us in sensible intuitions, hence necessarily representable only by reason. On the practical side, Kant holds that both freedom and the Categorical Imperative are Ideas of reason.[21]

Two features of this interpretation of Kant's broader account are particularly relevant to O'Neill's version of constructivism. First, as noted above,

one of O'Neill's main motivations for rejecting this sort of metaphysical account is that she thinks that the transcendence of its objects, such as God and freedom, precludes the possibility of agreement. In response, it is important to note that Kant does not think that such real objects are completely transcendent. It is true, in fact crucial to Kant's aims, that we cannot attain the same kind of knowledge of such objects that we can of empirical objects. However, even if the existence of such objects is not fully demonstrable, Kant does not maintain that they are therefore completely inaccessible to us. As we have seen, he defines reason in such a way that it can grasp these objects, even if he insists on the fact that we cannot demonstrate with theoretical certainty that the objects represented by reason exist precisely as they are represented. But as soon as we concede that it is the essential task of reason to grasp such real objects, no reason remains for thinking that agreement on them is necessarily impossible or, for that matter, any more difficult than agreement on principles that we might construct.

Second, it is clear that this interpretation of Kant's broader account of reason is intended to be understood as a version of realism rather than constructivism. On Kant's position, whatever is unconditioned is unconditioned, and our task as rational knowers and agents is to determine what the unconditioned demands of us by using the only faculty we have that can accomplish such a task, reason. One natural way of seeing the realist side of this interpretation of Kant's position is to note that his account of reason is not detachable from Transcendental Idealism insofar as appearances are always conditioned and subject-dependent given that they depend on our forms of intuition, whereas things in themselves, including God, our soul, and freedom, can be unconditioned and are not subject-dependent. They are things as they really are in themselves.[22] If we accept such a realist interpretation of God, freedom, and the Categorical Imperative, it may also put us in a position to give a fuller account of the sense of normativity that is typically ascribed to the practical principles of morality and justice, since the reality of independently existing entities can provide a more solid foundation for explaining a fuller sense of normativity than the modalities involved in O'Neill's version of constructivism.

## 4. Korsgaard on the Value of Humanity

In *Creating the Kingdom of Ends*, Christine Korsgaard poses the following question: "Does Kant think, or should a Kantian think, that human beings simply have unconditional or intrinsic value, or is there a sense in which we must confer value even upon ourselves?"[23] The answer she says she now favors is the sense in which we must confer value, and elsewhere she makes it clear that she thinks it is a mistake to suppose that "the value of persons [is] a metaphysical reality, perhaps in need of a metaphysical defense."[24]

Korsgaard's interpretation of Kant's or Kantian ethical thought is motivated in large part by the desire she shares with O'Neill to avoid the need for "ontological or metaphysical commitments" to back up claims of value.[25] On her interpretation, value and normativity are not to be met with or discovered in the world, but are generated and imposed on the world by the exercise of rational agency. Indeed, value and normativity are to be understood simply in terms of evaluative or normative commitments allegedly generated by the very functioning of the will, and they have no further metaphysical reality. As with O'Neill's approach, this avoids any appeal to metaphysical claims about value or normativity, which Korsgaard thinks are liable to run afoul of the modern scientific worldview or the metaphysics of the modern world, and are in any case explanatorily impotent.[26]

Korsgaard's approach to Kantian ethics may be described as a radically constructivist one, for two reasons. First, constructivism is applied comprehensively, not only to the value and normative force associated with our various subjective ends, but also to the value at the heart of Kant's moral philosophy, the intrinsic and unconditional value of humanity itself. It is constructivism all the way down. Second, the constructivism in question is broader than the familiar claim that facts about value are grounded in facts about the possibly hypothetical outcomes of possibly hypothetical procedures carried out by rational agents, and have no independent existence or grounding of the sort a realist might posit. That basic proceduralist claim by itself, which may be called weak constructivism, would still be consistent with facts about value existing independently of the actual carrying out of the procedures in question. They might exist, for example, as abstract facts about what rational agents would choose, or value, or be committed to if they were to deliberate, perhaps under certain specified conditions behind a veil of ignorance or with full information. On that picture, it might then be thought that the business of ethics is to discover those facts about value, through a theoretical investigation of the nature and structure of rational deliberation. Korsgaard, however, takes facts about value to be not merely dependent on the hypothetical outcomes of certain procedures in this way, but to be created in the first place only by the actual carrying out of the procedures in question, in particular, by certain procedures generically involved, of necessity, in the very functioning of the will. Evaluative and normative facts are not there as abstract facts to be met with or discovered through theoretical investigation of the nature and structure of rational agency, but are constructed through our actual practical activities, contrasting to both realism and weak constructivism.

This feature of Korsgaard's view becomes clear in her discussion of the status of moral facts in connection with the universal law formulation of the categorical imperative. Someone might be tempted to read Kant as holding that although moral facts are derived solely from a principle describing procedures necessary to the exercise of rational agency, and have no independ-

ent existence as on realist views, they nonetheless exist timelessly as abstract facts that are in some sense there to be discovered theoretically, as facts about the fitness or unfitness of a maxim to be a law in a kingdom of ends. Korsgaard, however, rejects such a weak constructivist interpretation in favor of what we have called a radical constructivist one, according to which: "values are created by human beings. Of course we can discover that the maxim is fit to be law; but the maxim isn't a law until we will it, and in that sense create the resulting value. . . . Values are constructed by a procedure, the procedure of making laws for ourselves."[27] On her view, values or normative facts are practical not merely in content but in nature, coming into being only with the actual exercise of rational agency. It is this radically practical approach to ethics that leads Korsgaard to suggest that ethics is not really a theoretical discipline at all. It is not a "branch of knowledge, knowledge of the normative part of the world" but something fundamentally practical in a way that excludes that conception.[28]

Let us return to Korsgaard's approach to the value of humanity, focusing on her interpretation of Kant's argument for it.[29] The argument is roughly as follows. Since we are reflective beings, our practical functioning essentially involves acting for reasons. But we can act for reasons only if we regard some considerations as genuinely good reasons for acting, which in turn involves regarding some ends as genuinely good. But in order to regard some ends as genuinely good, "we must regard ourselves as capable [by virtue of our rational nature or our capacity to set ends] of conferring value upon the objects of our choice, the ends that we set."[30] The only way we could have this value-conferring capacity *qua* rational agents would be for us to be unconditionally valuable *qua* rational agents. Hence, in order to function practically, we must regard ourselves as unconditionally valuable *qua* rational agents. This is how Korsgaard understands Kant's claim, as she puts it, "that regarding your existence as a rational being as an end in itself is 'a subjective principle of human action'."[31] If we must regard ourselves in this way, simply because of our possession of rational nature, then we are equally committed to so regarding other rational beings. Thus, insofar as we are to function practically at all, we are committed to regarding rational agents as unconditionally valuable ends in themselves. That is the sense in which others have unconditional value which has normative force for us. It is not some special fact about them which we are bound to recognize and respect but a matter of commitments we are already involved in as a result of certain generic conditions on the exercise of our own rational agency.

On this interpretation of the argument, the value of rational agents comes into being as a practical commitment on the part of the valuers who are also rational agents, and has no further reality than that. We are committed to regarding ourselves and others as valuable, much as Kant argues in the third section of the *Groundwork of Metaphysics of Morals* that we are committed

to regarding ourselves as free, and so we need not concern ourselves further about metaphysical facts, which have become irrelevant. One important difference, however, is that in the case of value it is not merely that we need not concern ourselves with the metaphysical facts. Korsgaard apparently wants to deny any metaphysical fact of value. All there is here is the practical commitment, which exhausts the value in question. Anything further would be undesirable realist baggage.

## 5. Problems for Radical Constructivism

Korsgaard thus takes Kant to be offering an account of the unconditional value of rational nature or humanity according to which it is decidedly not a kind of metaphysical fact, but consists in the practical commitments generated by certain generic necessary conditions on the exercise of the will.[32] One result of this is that the value of humanity turns out, in an important sense, not to be basic after all. It is not simply a fundamental property of humanity, as a realist would have it, but winds up being further explicated in terms of a certain modal fact about the nature of willing: the alleged fact that we cannot get our wills to function unless we think certain things about ourselves. That general modal fact about the nature of willing appears to be more fundamental than the very value that is supposed to lie at the heart of Kant's view, since it is what generates the alleged commitments in terms of which that value is being construed.

This feature of the view is less than attractive if we are inclined to suppose that the unconditional value of rational beings or rational nature is a fundamental starting point, and not something that can be explicated in terms of some much less exalted modal fact about the nature of willing. But Korsgaard will presumably defend it by claiming that it is the only way to account for the normative force of such value for us. If we are already committed to such value by the procedures we must follow in order to exercise our wills at all, as would clearly be the case if such value is just explicated in those terms, then there is no mystery about how such claims of value have normative force for our wills. By contrast, realists might seem to have a hard time here. They will claim, contrary to the earlier quote from Korsgaard, that it is a fact, to be recognized and respected by rational agents, that human and other rational beings simply have unconditional or intrinsic value by virtue of their possessing rational nature. But how will realists explain how that external fact binds the will of a deliberating agent who encounters it, so that he must recognize its normative force to constrain his actions?

In fact, however, there is good reason to question whether Korsgaard's radical constructivism contains an adequate account of normative force after all. As part of the practical reduction of the value of persons to the commit-

ments of deliberators arising from generic conditions on the exercise of rational agency, the radical constructivist approach involves an unattractive account of the source of other people's value insofar as it has normative force for us. On Korsgaard's interpretation of Kant's view, what stands in the way of my treating you as a mere means to some end of mine is not your metaphysical possession of intrinsic, unconditional value, which presents itself to me as a fact about you that I encounter and have to respect. Instead, it is the fact that I am committed, as part of what is generically involved in the exercise of my own will, to regarding you, *qua* possessor of rational agency, as having such value. Of course, it is also true that you confer such value on yourself, as I do on myself; each of us regards himself or herself as being valuable. But the way in which your value is supposed to have normative force for my will is that I am committed to regarding you as having it, recognizing that consistency requires this given my regarding myself as valuable simply because of my possession of a rational nature, which I must do in order to act at all. That is the way in which your value is supposed to get a grip on my will, rather than my simply being called upon to respond to your value as something presented to me from outside of my will, which is precisely what Korsgaard finds objectionable about realism, and so cannot be her interpretation of what is going on here.[33]

This, however, does not seem like a very satisfactory representation of the source of other people's value and of its normative force for us. What is wrong with enslaving someone, for example, seems to be something straightforwardly and simply about her, given what she is – the dignity that belongs to her as a rational being. To cash out the wrongness of such an action and its normative force for me in a way that requires a detour through a story about what I have to do in order to exercise my will at all seems like a move in precisely the wrong direction. It does not seem true to ordinary moral experience, which certainly does not represent other people's value and its significance for us as deriving from commitments bound up with the exercise of our own wills under certain generic constraints inherent to the nature of willing. The phenomenology, for what it is worth, is that other people, as rational agents, simply matter, and that this makes it inappropriate for us to treat them as if they did not, apart from any commitments that might arise generically through the exercise of our own wills.

It is revealing to note the similarity between this criticism of radical constructivism and a criticism that is often raised against non-cognitivism, and seems to apply even to sophisticated forms of it, such as Simon Blackburn's account of quasi-realist projectivism.[34] Suppose a projectivist shares the Kantian first-order moral judgment that rational agents have unconditional value, which places constraints on how they may be treated. He will nonetheless maintain with Korsgaard and against the realist that this is not a metaphysical fact about persons. A person's rational nature does not directly and

objectively ground her value or any practical norms. Instead, he will offer a projectivist meta-ethical account of such value and the norms it grounds that involves a detour through a discussion about non-cognitive attitudes on the part of valuers. This of course is importantly different from Korsgaard's cognitivist Kantian discussion. But it raises a parallel concern, which Blackburn has often recognized and attempted to answer: Is it really plausible to suppose that the explication of what is wrong with enslaving someone, for example, will involve an appeal to such facts about us and our attitudes? What is wrong with such behavior seems to have to do simply with something about the victim, the intrinsic value rooted in her rational nature, independently of such things as our attitudes toward slavery.

Blackburn's familiar reply is simply to shift to the level of first-order moral judgment whenever such concerns are raised at the meta-ethical level.[35] He notes, for example, that his first-order disapproval of slavery is based simply on its offending against the dignity of rational nature and not on anything about contingent human attitudes toward slavery. He does not, after all, cite our attitudes in explaining why slavery strikes him as objectionable. That enables him to give a certain sense to the claim that what makes it wrong to enslave someone is simply her dignity as a rational agent, and not anything about our attitudes, so that slavery would be wrong even if we all felt otherwise. By interpreting such claims as remarks about the precise shape of his first-order attitudes and their objects, rather than as meta-ethical claims, he is able to make realist sounding claims, right down to claiming a kind of mind-independence, for moral facts. This is unsatisfying, however, because despite his ability to say such things at the level of first-order moral judgment, there is no escaping the fact that at the meta-ethical level he is still forced by his account of projectivism to say that the value and normativity are ultimately dependent on the attitudes of appraisers. This is enough to be problematic. Why should we suppose that the proper meta-ethical account of the wrongness of something like slavery involves such a detour through a story about the attitudes of the appraisers any more than the explanation at the first-order level does? It seems far truer to our moral experience just to say, from both the normative and the meta-ethical perspectives, that the wrongness consists directly in the fact that the action egregiously fails to respect the dignity and value of the victim as a rational being whose nature directly makes such treatment by any moral agent capable of recognizing the value in question inappropriate. Something has gone awry if the theory requires us to look elsewhere, even if only at the meta-ethical level.[36]

Our claim, then, is that something parallel to this is going on in the radical constructivist interpretation of Kant. There seems to be a commitment at the meta-ethical level to a detour in the explication of another person's value insofar as it is normatively significant for me through a certain modal understanding about me or about what I have to be committed to in order to exer-

cise my will at all. This may preserve a guarantee of other people's value for me, unlike with sentiment-based views, since it is not dependent on anything contingent about me, but it still seems like a very odd sort of thing to look to as its source.

The same concern can be raised about O'Neill's radical constructivist account of the closely related matter of the scope of moral consideration, of who counts as having moral standing, which should constrain our actions. O'Neill does not take the question of the scope of moral consideration to be theoretical, with an answer waiting to be discovered through metaphysical argument. Instead, she thinks it is fundamentally a practical problem to be solved by giving a usable procedure for answering the questions as they arise, suitable for practical purposes.[37] Her position may be summarized by saying that "conclusions about the scope of ethical consideration are derived from the assumptions to which agents commit themselves in acting."[38] The idea is that our actions themselves are often predicated on certain assumptions about other people's agency, our possible effects on them or *vice versa*, and their vulnerabilities. When we perform such actions, we thus commit ourselves to those assumptions, and that means that consistency requires recognizing the same assumptions when thinking about other people's moral status:

> When agents commit themselves to the assumption that there are certain others, who are agents or subjects with these or those capacities, capabilities and vulnerabilities, they cannot coherently deny those assumptions in working out the scope of ethical consideration to which they are committed. Commitments to others' ethical standing are taken on as soon as activity is planned or begun.[39]

For present purposes we can set aside the question whether such an approach will yield sufficiently robust and pervasive commitments to give plausible answers to questions about the scope of moral consideration. The point to emphasize is just the similarity with Korsgaard's radical constructivism. The moral standing of other people is not a metaphysical fact about them to be recognized and respected by me, but is to be explicated in terms of alleged commitments on my part growing out of my own practical activities. In particular, on O'Neill's view, your having a moral standing that makes claims on me just consists in my being committed to regarding you in a certain way as a result of certain assumptions implied by my own actions. There is the same sort of practical reduction of the value, shifting away from the object of value itself and appealing to certain involvements of potential valuers as a result of their activity. Again, we might wonder whether such an account of the nature and source of other people's value is adequate to our experience of the value of persons and of the kind of normative force it seems to have for us.

## 6. An Alternative

The natural alternative, already mentioned, would be to take other people's value simply to be a fact about them, rooted directly in their rational nature, which we are able to recognize and are therefore obliged to respect simply because we are capable of recognizing it. There is no reason to suppose that value must be practically reduced to our own commitments in order to have normative force or authority for us, though of course how we actually respond to it is another question. We can move in a more realist direction while remaining faithful to the Kantian focus on the value of humanity as the irreducible value lying at the heart of ethics. It is hardly obvious why such a limited move toward realism should be thought somehow inconsistent with modern science or metaphysically extravagant. After all, it involves no suggestion that value and normativity exist independently of rational nature, out there in the non-rational world; in fact, the present suggestion does not involve positing anything transcendent or outside of space and time, such as God, and so it could be accepted even by someone who rejected the earlier appeal to Kant's Ideas of reason and their noumenal objects. It is simply a matter of taking the value of rational beings to be a non-constructed fact about them.

Korsgaard is perhaps correct in suggesting that realism would not really explain our "confidence in our claims of value," but is rather something we are led to do, if we are, because of our confidence in such claims.[40] But that is not an objection. The point is that construing other people's value as a straightforward fact about them simply accords more deeply with our moral experience than the radical constructivist accounts, which instead seem to locate value and normativity in the wrong place and to ground them in the wrong sorts of things. If that is right, and if we have not been given compelling reason from modern science for doubting the moral appearances, then we have sufficient reason for adopting a more realist position, whether or not this figures into any further explanatory discussion.

If Kant did not intend to be offering a practically reductive account of the value of humanity in the argument discussed earlier, why did he emphasize, as he clearly did, our universal commitments as rational agents to the value of humanity? The argument can be seen as an attempt to show that this is a value we are in any case already committed to, so that for practical purposes we need not concern ourselves with a metaphysical defense of it. The categorical imperative it grounds is in any case a valid law for us. There is certainly a point to that sort of argument, even if we also think that there are metaphysical facts here as well, which again, would be very much along the lines of the argument concerning freedom in the *Groundwork of the Metaphysics of Morals.* It might still be important to posit such facts about value for the simple reason that the alternative, radical constructivist construal of such value fails to do justice to our considered experience of it.

An example of a more realist approach to Kant may be found in Allen Wood's recent work, where it is claimed that "Kant's realism precludes identifying moral truth with what is 'constructed' through any (Rawlsian) 'CI-procedure'."[41] Instead of taking Kant to be offering an argument to demonstrate the value of humanity by reducing that value to radically constructivist terms, he takes Kant to hold that claims of ultimate value are "indemonstrable."[42] The point of the argument, he suggests, is similar to that of J.S. Mill's proof of the admittedly indemonstrable principle of utility. On this approach, a principle of ultimate value can be

> argued for rationally (and even 'proven' in a looser sense of 'proof'), by showing that what the principle takes to be valuable is already 'in theory and in practice, acknowledged to be an end'. Any argument of this kind must begin with value judgments we already accept and then provide a convincing theoretical interpretation of these judgments that supports the pertinent philosophical claim about ultimate value.[43]

In Kant's case, the idea is that the argument will "show that we already do (and that we must) value" rational nature as an end-in-itself, and by doing so, it will help "to convince us that rational nature has such value."[44] On this reading, Kant holds that it is simply an indemonstrable but universally recognized fact that rational nature has the value in question. The significance, then, of our universally taking rational nature to have such value is not that this reveals where the value comes from or what it consists in, as on the radical constructivist view, but that it allegedly helps to convince us that such value is not a mere illusion, but is indeed a real feature of rational nature that we are already committed to recognizing.

## 7. Conclusion

It would require a great deal of textual work to determine which of these approaches provides the more faithful reading of Kant. Our concern here has been simply to bring out what seem to be significant drawbacks to the constructivist approach to ethics represented in the prominent work of O'Neill and Korsgaard. A more realist approach to the value at the heart of Kant's moral philosophy, and to the issue of normativity in general, remains an attractive alternative that can avoid the potential difficulties of both radical constructivism and more extreme forms of realism, while remaining true to Kant's emphasis on humanity as the source of value and normativity.[45]

## Notes

1. See Christine Korsgaard, *The Sources of Normativity* (Cambridge, England: Cambridge University Press, 1996), pp. 5 & 37, and Onora O'Neill, *Towards Justice and Virtue* (Cambridge, England: Cambridge University Press, 1996), pp. 3, 6, 51, & 191–197. See also J.B. Schneewind's *The Invention of Autonomy: A History of Modern Moral Philosophy* (Cambridge, England: Cambridge University Press, 1998).
2. See also Karl Ameriks, "On Schneewind and Kant's Method in Ethics," *Ideas y Valores* 102 (1996), pp. 28–53.
3. John Rawls, "Kantian Constructivism in Moral Theory," *Journal of Philosophy* 77 (1980), p. 516.
4. Ibid., p. 517. See also Thomas Hill, "Kantian Constructivism in Ethics," *Dignity and Practical Reason in Kant's Ethics* (Ithaca, N.Y.: Cornell University Press, 1992), pp. 226–250.
5. Rawls, op. cit., p. 519.
6. O'Neill, *Constructions of Reason: Explorations of Kant's Practical Philosophy* (Cambridge, England: Cambridge University Press, 1989), p. 210.
7. Rawls, op. cit., p. 518.
8. O'Neill, *Constructions of Reason*, p. 211.
9. See Patrick Kain, "Self-Legislation and Prudence in Kant's Moral Philosophy: A Critical Examination of Some Constructivist Interpretations," ch. 3, Ph.D. thesis, University of Notre Dame, November, 1999.
10. O'Neill, *Constructions of Reason*, p. 210.
11. See ibid., p. 212.
12. Ibid., p. 213. See also O'Neill, *Towards Justice and Virtue*, p. 52.
13. O'Neill, *Constructions of Reason*, p. 215.
14. Cf. O'Neill, *Towards Justice and Virtue*, p. 48.
15. O'Neill, *Constructions of Reason*, pp. 46–47.
16. Ibid., pp. 19–20.
17. Ibid., p. 35.
18. Ibid., p. 38.
19. Ibid.
20. O'Neill, *Towards Justice and Virtue*, p. 60.
21. See Allen Wood, *Kant's Ethical Thought* (Cambridge, England: Cambridge University Press, 1999), pp. 157–158.
22. See Henry Allison, *Kant's Theory of Freedom* (New York: Cambridge University Press, 1990). See also Allen Wood, *Kant's Ethical Thought*.
23. Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge, England: Cambridge University Press, 1996), p. 407.
24. Ibid., pp. 407 & 301.
25. Korsgaard, "The Reasons We Can Share," *Creating the Kingdom of Ends*, p. 305n17.
26. Ibid., and Korsgaard, *The Sources of Normativity*, pp. 5 & 37.
27. Korsgaard, *The Sources of Normativity*, p. 112.
28. Ibid., pp. 37, 44–47, 89 & 254.
29. See Korsgaard, "Kant's Formula of Humanity,"and "Two Distinctions in Goodness," *Creating the Kingdom of Ends*, pp. 260–261. See also Korsgaard, *The Sources of Normativity*, p. 122.
30. Korsgaard, "Two Distinctions in Goodness," *Creating the Kingdom of Ends*, p. 260.
31. Ibid.
32. Cf. Korsgaard, *The Sources of Normativity*, pp. 89 & 254.

33. See Korsgaard, "The Normativity of Instrumental Reason," *Ethics and Practical Reason*, eds. Garrett Cullity and Berys Gaut (Oxford: Oxford University Press, 1997), pp. 219, 240 f.
34. See Simon Blackburn, *Ruling Passions* (Oxford: Oxford University Press, 1998), and Blackburn, *Spreading the Word* (Oxford: Oxford University Press, 1984).
35. See Blackburn, *Spreading the Word*, p. 218; "Errors and the Phenomenology of Value," in Blackburn, *Essays on Quasi-Realism* (Oxford: Oxford University Press, 1993), p. 153; and Blackburn, *Ruling Passions*, p. 311.
36. See Thomas Nagel, *The Last Word* (Oxford: Oxford University Press, 1997).
37. O'Neill, *Towards Justice and Virtue*, pp. 93–97.
38. Ibid., p. 110.
39. Ibid., p. 100.
40. Korsgaard, "The Reasons We Can Share," p. 305n17, and Korsgaard, *The Sources of Normativity*, pp. 39–41.
41. Allen Wood, op. cit., p. 374.
42. Ibid., p. 125.
43. Ibid.
44. Ibid.
45. We would like to thank Karl Ameriks, Patrick Kain, and especially Patrick Croskery for helpful comments on earlier versions of the first three sections of this paper.