The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity*

William J. FitzPatrick

Christine Korsgaard has put forward a sustained critique of ethical realism as part of a larger argument for reconceiving the nature of practical philosophy itself.[1] What turns out to be most problematic about realism, in her view, is not just the familiar set of metaphysical and epistemological worries others have pressed but the very fact that "realism conceives ethics on the model of applied knowledge."[2] This generic feature of realism—whether naturalistic or nonnaturalistic, reductionist or nonreductionist, Platonic or mundane—allegedly embodies a failure to appreciate the deeply practical nature of normativity, making realism incapable of accounting for the normative force ethical requirements have for us. The only solution, according to Korsgaard, is to dispense with realism once and for all and to embrace a form of neo-Kantian constructivism that conceives ethics as the "work[ing] out [of] practical solutions to practical problems" in a way that excludes any realist notion that there is "some sort of ethical knowl-

edge that we can apply in action."[3] It is this call for a fundamental practical turn in ethical theory that is at the heart of her larger project.

My aim is to assess the merits of Korsgaard's rejection of realism as well as the prospects for her practical approach to normativity, and I shall both raise problems for her constructivism and develop a realist response to her central challenge. To make clear what is at stake, I will begin by elucidating the "normative question" that motivates her negative and positive projects alike and by describing the realist and constructivist positions in this debate. Then, in Section II, I will lay out her central critique of realism and go on, in Section III, to explain her current strategy for avoiding the problems she raises for realists. As described more fully at the end of Section III, I will then examine Korsgaard's distinctive approach to normativity by looking at how her practical-problem-solving conception of normativity applies specifically to the principle at the heart of Kantian ethics—the formula of humanity, which she believes must be derived in a purely constructivist fashion in order to be binding. Giving careful attention to such derivations in light of her general account of normativity (Secs. IV–VI) is the best way both to clarify what is really at issue in the debate over normativity and to see concretely how her constructivist approach to deriving normative principles is supposed to deliver what is required by her practical-problem-solving account of binding normative force.

I shall argue that such an examination reveals crucial ambiguities and holes in the constructivist derivations of the formula of humanity, raising questions about whether and how they can deliver what the broader conception of normativity requires. After exploring the problems that arise for Korsgaard here, particularly in light of her own arguments against realism, I will consider in Section VII how she might remedy one central difficulty, retaining a weaker version of her practical-problem-solving conception of normativity. The problem, however, is that this approach would no longer enjoy the alleged advantages over realism. Indeed, in developing a realist answer to her central challenge, in Sections VIII and IX, I will argue that realists can co-opt the very same practical-problem-solving model of normativity and employ it within a realist framework. Korsgaard's intriguing focus on a principle's role in solving an unavoidable practical problem involving agency, in order to secure its normative relevance to the will, turns out to be divorceable from a purely constructivist, practical account of the *grounds of the solution* to that problem. I will show how a realist can thus exploit the former and close any gap between normative principles and the will without abandoning a realist construal of those principles themselves.

My conclusion will be therefore that we have not been given good

3. Ibid., 118.

reason either to reject realism as an antiquated relic of the last millennium or to think that we as yet have a viable alternative in a purely constructivist view that construes ethics as practical "all the way down." Constructivism may indeed have some role to play in ethics and perhaps a large role in political philosophy, given the special nature of the problems characteristic of the political sphere. The question is whether it must, or even can, play the sort of general and pervasive role in ethics that Korsgaard envisions in recasting ethics as a purely practical discipline.

I. THE NORMATIVE QUESTION AND THE REALIST-CONSTRUCTIVIST DIVIDE

A central question about moral requirements is whether they, unlike the arbitrary requirements of a system of norms such as etiquette, are inescapable in the sense that they provide genuine reasons for acting regardless of an agent's contingent desires and interests.[4] This question about morality is at the heart of the traditional search for a "justification for morality" or an account of the "rational authority of morality" or of the genuine "normativity" or "reason-giving force" of morality.[5] What is at issue here is "whether considerations that present themselves as reasons from within the moral point of view . . . *really are reasons*"[6]—not just moral reasons (trivially) but genuine reasons concerning "what the agent rationally *ought* to do."[7] This is important because, as Darwall has stressed, "the rational 'ought' is almost never treated by philosophers as simply one 'ought' among others, on all fours with 'oughts' *internal* to e.g. etiquette, baseball or bridge. It is regarded, if only implicitly, as *unqualifiedly normative*—not just an ought-according-to-the-norms-of-rationality, as there might be oughts-according-to-the-norms-of-etiquette, or -baseball, or -bridge. What a person rationally ought to do is whatever he ought to do *simpliciter*—*sans phrase*, as it were." The question, then, is whether moral requirements enjoy unqualified normative authority, bearing directly on what any person most fundamentally ought to do.

This is in part a theoretical question about how morality fits into

4. See Philippa Foot, "Morality as a System of Hypothetical Imperatives," in her *Virtues and Vices* (Berkeley: University of California Press, 1978).

5. The quoted phrases are taken, respectively, from Stephen Darwall, "Autonomist Internalism and the Justification of Morals," *Noûs* 24 (1990): 257–68; David O. Brink, "A Puzzle about the Rational Authority of Morality," in *Philosophical Perspectives* 6, ed. James E. Tomberlin, 1–26 (Atascadero, CA: Ridgeview, 1992); Korsgaard, *The Sources of Normativity*; and Foot, "Morality as a System."

6. Darwall, "Autonomist Internalism," 258.

7. Stephen Darwall, "Internalism and Agency," in Tomberlin, *Philosophical Perspectives* 6, 155–74, 156.

a broader normative scheme, but as Korsgaard has stressed, it is more than that. The question of the genuine normative force of morality—which we may follow her in calling "the normative question"—is one that arises most powerfully from the perspective of first-person practical deliberation, particularly in the face of an onerous moral burden, and it is here that it becomes most clear why it matters. I may grasp that I am under a moral requirement to tell the truth about something I have done and yet also know that this will result in something I deeply wish to avoid. In this situation, I am confronted in a real, practical way with the question of the normative authority of morality: without denying the legitimacy of the content of the requirement from a moral point of view, I find myself with a strong incentive to deny its normative force—to set it aside in my deliberation, in favor of avoiding the sacrifice it demands. And why shouldn't I? What reason do I have—speaking again from the perspective of unqualified normativity—to honor this requirement despite this unwelcome cost, even if I could get away with violating it? A satisfactory answer to the normative question must be one that someone in this situation could reasonably be expected to find compelling.

Korsgaard has placed the normative question at the heart of both her own neo-Kantian constructivism and her attack on ethical realism. While there are many different versions of ethical realism, with widely varying claims of independence for ethical truths or properties, the target of Korsgaard's attack is the realist's commitment to the existence of ethical truths that are not themselves constructed from our use of ethical concepts but are to be discovered and represented by propositions that accurately employ those concepts.[8] According to realists, there are facts about what is good or should be done which are there to be recognized and expressed using our ethical concepts rather than being constructed through our constrained use of such concepts. The thought, for example, "that human beings simply have unconditional or intrinsic value" is a realist one standing in contrast to the constructivist claim that "we must confer value even upon ourselves" through certain allegedly necessary uses of ethical concepts.[9]

The constructivism Korsgaard embraces and contrasts with realism is a form of proceduralism according to which the rightness of correct answers to normative questions is grounded in the fact that these answers

---

8. Korsgaard, "Realism and Constructivism," 100, and 119 n. 3.

9. Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), 407. Korsgaard has occasionally used the expression 'substantive realism' to refer to what I am here calling 'realism', while referring to her neo-Kantian constructivism as 'procedural realism' (*The Sources of Normativity*, 36). For simplicity and clarity, I will stick with the labels 'realism' and 'constructivism', reflecting her more recent terminology.

are yielded by principles deriving from procedures with some special status. The now classic paradigm of such an approach is Rawlsian constructivism in political theory, wherein answers to questions about justice count as correct simply *because* they are yielded by principles deriving from a certain procedure that embodies the appropriate conditions of fairness and equality.[10] For Korsgaard, the relevant procedures at the source of normativity are procedures involved with willing, and what gives them their special status is that they are practically necessary for us—formal procedures rational beings *must* employ simply to function as agents at all. Everything starts with the nature of the will and the procedures according to which it must operate if it is to function as a will at all. These necessary procedures are then the basis from which principles that answer questions about morality and about reasons are derived, and the correctness of those answers lies in their being derived from such practically necessary procedures, rather than in their expressing prior facts.[11]

II. KORSGAARD'S CRITIQUE OF REALISM

What, then, is the central difficulty for realism in connection with the normative question? At times Korsgaard has claimed that realism "refuses to answer the normative question. It is a way of saying that it cannot be done. Or rather, more commonly, it is a way of saying that it need not be done."[12] This, however, is a mischaracterization of realism and fails to capture what is at issue. The realist's positing of non-procedure-based truths about the goodness or badness of actions and about reasons for acting in no way precludes taking the normative question seriously and attempting to answer it. Plato is a clear example of a realist who did just that. And when addressing an agent faced with an onerous moral requirement, for whom the normative question is pressing, realists are hardly limited to repeating pointlessly that "it is *true*" that there is such a requirement and that the agent has a reason to comply with it.[13] At least some realists will instead offer an account of the connection between moral facts and reasons for acting that shows why the moral requirement implies a genuine reason for so acting—a reason with unqualified normative force—despite the unwelcome costs. Whether this can in principle succeed or is somehow incapable of delivering what a satisfactory theory of normativity should give us is a deep question with which we shall be largely concerned in what follows. But it is important

10. See John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971).

11. See Korsgaard, *The Sources of Normativity*, 36, 120–23.

12. Ibid., 39.

13. Ibid., 38.

to be clear that realism implies neither a refusal to answer the normative question nor the view that this needn't be done.

Similarly, it is a mistake to claim that realism fails to address the normative question on the grounds that it presupposes the very confidence that is missing when the normative question is most pressing.[14] Realists do often fall back on a confidence in the existence of facts about morality or about reasons—as Korsgaard notes about Nagel. But that confidence should not be confused with a confidence in a positive answer to the normative question, which is about the *connection* between morality and reasons for acting. Someone may be confident that moral facts exist and that reasons for acting exist without being confident that moral facts automatically provide reasons for acting—as illustrated by Foot at the time she likened morality to etiquette with respect to reason-giving force. Thus, a realist might lack confidence in a positive answer to the normative question and yet still seek to develop such an answer by trying to bring out the relevant connection between moral facts and facts about reasons. And even where a realist already has confidence in the normative force of morality, he can still feel the need to try to answer the normative question: for although it arises most powerfully "when our confidence has been shaken," it can be raised just as cogently in a cool hour, when our confidence is strong and we merely seek a better understanding of the normative force of morality.[15] The real question is just whether there is something about realism that makes it essentially ill suited to answer the normative question in a satisfactory way. Here is where Korsgaard develops a more challenging argument rooted in a concern about how ethics is supposed to influence action.

On a realist view, the goodness or badness of an action is a fact about it to be recognized, and this knowledge is then typically to be applied in deliberation. But in order to apply such a piece of knowledge and come to be motivated, Korsgaard argues, one would need another norm dictating, for example, that one ought (or is obligated) to perform good actions. Where, then, is that norm to come from? It cannot itself be derived from the above knowledge of the goodness of an action, for the very question is why one should care about such goodness and apply such knowledge. Indeed, nothing we could know about the action—no knowledge of any feature of it construed in a realist fashion—could be applied practically without the aid of some further norm prescribing the application of such knowledge; and that further norm cannot itself be understood on the realist model, as a known normative truth to be applied, without just raising the same problem again. So it seems impossible to explain normative force in realist terms: if the will is to be

14. Ibid., 40–42.
15. Ibid., 40.

normatively bound to perform some action, this can be understood only in terms of some nonarbitrary norm for which this question of application does not arise (with the consequent need for further norms ad infinitum), as it will for any norm construed as a realist normative truth calling out to be recognized and applied in deliberation.[16]

This argument, Korsgaard says, may be thought of as "the ultimate extension of the open question argument: If it is just a fact that a certain action would be good, a fact that you might or might not apply in deliberation, then it seems to be an open question whether you *should* apply it."[17] If we then say that you should, this seems to require defending with another norm; and if that norm is itself taken to be a piece of knowledge (on a realist model, implying the existence of a normative fact that is known), then the same problem arises again with respect to it: why should you apply that piece of knowledge in your deliberation, especially if this would lead to some onerous sacrifice you may not be anxious to make? And so it continues, as long as we remain within a realist framework: it is a mystery how such alleged normative facts could have genuine normative force, binding your will.[18]

III. KORSGAARD'S CONSTRUCTIVIST PROGRAM

Korsgaard's solution is to abandon the realist framework and to construe normative force instead in such a way that these questions or normative gaps do not arise. The crucial proposal is to understand normative force in terms of practical problem solving.[19] We begin with a practical problem that is already a real concern for the agent as an agent. Such a starting point immediately eliminates the "why should I care?" question: from the first-person perspective of a deliberating agent, from which real practical questions about normativity arise, the concern with agency is given, and so therefore is the concern with practical problems concerning that agency. The project is then to derive a solution to such a

16. See Korsgaard, "Realism and Constructivism," 110–12, for this argument "against the model of application." Compare "Self-Constitution," lecture 2: "Practical Reason and the Unity of the Will," 13–17 (all page references to this forthcoming work are to the manuscripts of the original lectures). The argument is built off of a more specific argument against realism about instrumental normativity, also discussed in Korsgaard, "The Normativity of Instrumental Reason."

17. Korsgaard, "Realism and Constructivism," 112.

18. The difficulty here is related to a familiar problem for divine command theories. If we attempt to ground moral obligations in divine commands, the question arises: why are we obligated to obey divine commands in the first place? It will not help in answering this question merely to add more divine commands, including the second-order command to obey divine commands, or to argue for their genuine reality.

19. Korsgaard, "Realism and Constructivism," 112–19, and *The Sources of Normativity*, 113–14.

problem in a way that allows the necessary practical concern with the problem to carry over to the solution.

According to Korsgaard, this is to be done not by looking for some normative truth waiting to be discovered and applied but by constructing a practical solution to the problem at hand based simply on an adequate conception of that problem. So, for example, Rawls's principles of justice are supposed to be nothing more or less than a constructed practical solution to the fundamental practical problem faced by agents in a pluralistic liberal society, of finding principles of justice that are acceptable to all without compromising anyone's freedom. The principles constituting the solution are already implicit in an adequate conception of the problem, being arrived at "just by reflecting on the nature of that problem," and they constitute the solution because it is by employing these principles that agents are able to achieve the practical end in question—in this case living together on fair political terms as free and equal agents.[20] By thus starting with a real practical problem that must be dealt with, the constructivist approach neatly eliminates questions about why agents should care about the principles that are arrived at as solutions—which was precisely the difficulty that seemed to plague the realist with regard to his principles, conceived simply as normative truths to be applied (or not). For the constructivist, ethical or political principles are the solutions to practical problems we are already stuck with, and this is how normative force is explained: "If you recognize the problem to be real, to be yours, to be one you have to solve, and the solution to be the only or the best one, then the solution is binding upon you."[21]

Korsgaard's own approach in ethics focuses on Kant's formula of humanity, building on his attempt (as she sees it) to derive this fundamental principle of morality in a purely constructivist fashion, starting from a generic practical problem of agency—namely, an agent's need for reasons in order to exercise agency—and arguing for the necessity of adopting and conforming to the formula of humanity in solving it.[22] I shall examine these arguments carefully in what follows, as they constitute the clearest and most important applications of the practical-problem-solving model of normativity to Korsgaard's constructivist approach to moral principles. If successful, Korsgaard's arguments would establish the normative force of such principles for all agents by re-

20. Korsgaard, "Realism and Constructivism," 116.
21. Ibid. Compare Korsgaard, *The Sources of Normativity*, 114.
22. Kant's formula of humanity directs agents to "act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means" (Immanuel Kant, *Grounding for the Metaphysics of Morals*, trans. James W. Ellington [Indianapolis, IN: Hackett, 1993], Ak. 429).

vealing the practical role these principles play in the exercise of agency as such, and hence ultimately in our identity and existence as agents. As Korsgaard has recently put it, "The principles of practical reason serve to unify and constitute us as agents, and that is why they are normative": for "the necessity of conforming to the principles of practical reason comes down to the necessity of being a unified agent . . . [which] comes down to the necessity of being an agent . . . [which in turn] comes down to the necessity of acting . . . [which] is our plight. The principles of practical reason are normative for us, then, simply because we must act."[23] By contrast with realist views, there is no gap here between an agent and the principles that are claimed to have normative force for her. Thus, for Korsgaard, only this approach is able to answer the normative question: from the first-person perspective within which the normative question arises with real practical impact, the concern with agency is a given, and the principles are bound up with the very exercise of agency as such.

There are, however, really two parts to Korsgaard's proposal that need to be distinguished, as this affects the range of theoretical possibilities we recognize. On one hand, there is her claim that something's having normative force for an agent has to do with its being the solution to a practical problem the agent faces, thus eliminating the normative gap that allegedly plagues the realist. I will call this her practical-problem-solving conception of normative force. On the other hand, there is the claim that such practical problems are to be solved by "constructing" a solution simply on the basis of an adequate conception of the problem. I will call this her constructivist thesis. Having made this distinction, we may raise two fundamental questions that go to the core of both her attack on realism and her positive project.

First, is the practical-problem-solving conception of normative force really wedded essentially to the constructivist account of how solutions to such problems are to be arrived at? Or can they be pried apart, so that even a realist, who rejects the latter, might still exploit some version of the former in answering the objection about normative gaps? Second, with regard to Korsgaard's constructivist thesis: is her constructivist derivation of the formula of humanity really adequate to the task set by her own practical-problem-solving conception of normativity?

I shall begin with the second question and argue (in Secs. IV–VI) that such constructivist derivations—both Korsgaard's interpretation of Kant's argument and her own new version—run into problems that become especially clear when viewed in light of the practical-problem-solving conception of normativity. Even on the most charitable inter-

23. Korsgaard, "Self-Constitution," lecture 1: "The Metaphysical Foundations of Normativity," 19, 17.

pretation, the arguments fail to close the sort of normative gap Korsgaard has claimed undermines realism and are in fact vulnerable to the very same kind of "open question argument" she was trying to avoid. In Section VII, I will suggest a way in which she could modify her conception of how the practical-problem-solving model applies to her derivations, yielding a more plausible version of her project. This is of limited help to her, however, which brings us to the first question above. For the cost of such a move, focusing on "best" solutions to practical problems rather than on practically necessary ones, is the loss of the alleged advantage over realism. As I will argue in Sections VIII and IX, even if a principle's being the best solution to a practical problem faced by an agent is essential to its having binding normative force for that agent (and many realists may dispute this[24]), we can separate this from the idea that such a problem must be solved by "constructing" a solution simply from an adequate conception of the problem. Realists can make use of the former idea to anchor normativity—guaranteeing that a normative principle gets under one's skin, making a legitimate claim on the will—without abandoning realism about moral goodness and badness, thus answering Korsgaard's central objection.

IV. KORSGAARD'S RECONSTRUCTION OF KANT'S DERIVATION OF THE FORMULA OF HUMANITY

Korsgaard describes her own argument for the formula of humanity as "just a fancy new model" of Kant's argument.[25] The best way to understand her argument is therefore first to look carefully at her interpretation of Kant's argument, which directly shapes her own. In fact, the central difficulties with Korsgaard's new version are already present in her formulation of Kant's argument, as we shall see.[26]

24. A realist might argue, e.g., that he has the resources for bridging the alleged normative gap by arguing contra Korsgaard "that the goodness of action is directly normative for the agent" ("Realism and Constructivism," 112). For steps in this direction, see Derek Parfit, "Reasons and Motivation," *Proceedings of the Aristotelian Society* 77, suppl. (1997): 99–130. If this could be made out, we would not ultimately need to appeal to the idea of solving a practical problem in order to understand the idea of normative force.

25. Korsgaard, *The Sources of Normativity*, 122.

26. I take no position on the correctness of Korsgaard's Rawlsian-constructivist interpretation of Kant. Such an approach to Kant has been influential, though some dispute it, especially in connection with the question of basic human value. For an alternative view, see Allen Wood, *Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999), esp. 374 n. 4. Paul Guyer also seems to take more of a realist approach to Kant, in speaking of our "immediate recognition" of "the fundamental normative fact . . . that freedom has an 'inner value, i.e. dignity'" or "the normative recognition that this free will has an incomparable value," which sounds realist rather than constructivist in Korsgaard's sense (Paul Guyer, "The Value of Reason and the Value of Freedom," *Ethics* 109 [1998]: 22–35, 32–33, 30, 34).

Korsgaard emphasizes that Kant's argument proceeds "from the standpoint of the deliberating rational agent," and the reason why is clear from what has already been said: on her view, we can establish the normative force of moral requirements or principles for an agent only by seeing them as part of a solution to a practical problem faced by the agent. The practical problem Korsgaard fixes on is one pertaining to rational agency as such: the agent's need for reasons. For "our plight as self-conscious beings is that we find ourselves faced with the necessity of making choices and so in need of reasons to act." Self-consciousness sets up a "space of reflective distance from our incentives" such that "the issue of *reasons* arises."[27] Rather than being governed by instinct, we can ask of any given inclination whether or not it constitutes a good reason for acting; indeed, not only can we pose such questions but we cannot escape them: we need reasons for acting—considerations we regard as good reasons "to do one thing rather than another"—if we are to act at all, given the reflective structure of our consciousness.[28] The argument thus begins with this fundamental problem of agency and proceeds to draw out various alleged conditions for its solution and perhaps implications of those conditions, leading ultimately to the formula of humanity itself.

Although Korsgaard has discussed this argument in a variety of contexts over many years, she has consistently employed language that is ambiguous in crucial respects and has never laid out the argument with sufficient precision to make its structure fully clear, all of which makes for well-known difficulties of interpretation. In attempting the most charitable reconstruction of her reconstruction of Kant's argument, I shall draw both on a variety of texts where she discusses it explicitly and, equally importantly, on her more general discussion of normativity and practical problem solving. The latter is obviously important insofar as the argument must reasonably satisfy the conditions of the practical-problem-solving conception of normativity if it is to be successful by Korsgaard's own lights in establishing the normative force of the formula of humanity and of the moral requirements it yields. The issues here are complex enough, however, that I shall begin by setting the central interpretive difficulties to one side and laying out the argument first in a general form, sticking closely to the relevant texts.[29]

27. Christine Korsgaard, "Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind," *Ethics* 109 (1998): 49–66, 62.

28. Ibid., 51. See also Korsgaard, *The Sources of Normativity*, 121–22, 252–53.

29. In particular, Korsgaard, *The Sources of Normativity*, 122; Christine Korsgaard, "Kant's Formula of Humanity," in her *Creating the Kingdom of Ends*, 106–32, 121–23, and "Two Distinctions in Goodness," also in *Creating the Kingdom of Ends*, 249–74, 260; Korsgaard,

The structure of the argument is most cogently brought out by seeing it as a chain of claims about what any agent as such *must* do (in a sense to be clarified, for various steps, in the next section), beginning with the practical problem of agency and concluding with the normative principle to be established, showing it to be in some sense necessary as part of the solution to that problem, which we must solve if we are to function as agents at all:[30]

1. In order to function as agents at all (which we cannot avoid, as self-conscious beings), we *must* act on reasons, which means that we *must* regard some considerations as providing us with good reasons to act in certain ways rather than in others.

2. In order to regard some considerations as providing us with good reasons to act in certain ways, we *must* regard some ends as being genuinely good, that is, worth pursuing.[31]

3. If we are to regard some ends as genuinely good, "we *must* regard ourselves as capable [by virtue of our rational nature] of conferring value upon the objects of our choice, the ends that we set," that is, we "*must* see ourselves [qua possessors of rational nature] as value-conferring" and "regard [*ourselves*] as the source of [the] value" of those ends.[32]

4. If we are so to regard ourselves, we "must . . . take ourselves to be important" or "set a value on ourselves" and indeed *must* regard ourselves as unconditionally valuable qua possessors of rational nature.[33]

5. Therefore, we *must* regard ourselves as unconditionally valuable qua possessors of rational nature (1–4).[34]

6. And if we are to regard ourselves as having such value simply by

---

"Motivation, Metaphysics," and the crucial general discussion of normativity in "Realism and Constructivism."

30. This view of normative principles as constructed solutions to practical problems of agency is most clearly and explicitly articulated in Korsgaard, "Realism and Constructivism."

31. See Korsgaard, *The Sources of Normativity*, 122: "[Kant] started from the fact that when we make a choice we must regard its object as good."

32. Korsgaard, "Two Distinctions," 260, and "Motivation, Metaphysics," 60–61, 62 (emphasis added, except for the last). Compare also Korsgaard, "Kant's Formula of Humanity," 123: "In our actions we view ourselves as having a value-conferring status in virtue of our rational nature." This crucial third premise will be discussed in detail below.

33. Korsgaard, *The Sources of Normativity*, 122, and "Motivation, Metaphysics," 64. Compare the latter, 60–61: "Human beings must see ourselves as value-conferring and must therefore value humanity as an end in itself," and "Kant's Formula of Humanity," 123: "To play this role [i.e., as the source of our value-conferring ability] . . . rational nature must itself be something of unconditional value."

34. 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'" ("Two Distinctions," 260).

virtue of our possession of rational nature, we *must* so regard other beings equally possessed of rational nature.[35]

7. Therefore, we *must* regard rational beings as unconditionally valuable ends in themselves, which is the essence of the formula of humanity (5–6).

This understanding of Kant's argument is shaped both by Korsgaard's constructivist reading of Kant and by her view that Kant correctly locates normativity ultimately in necessary solutions to generic practical problems of agency, as set forth in the first premise. On this view, the formula of humanity has binding normative force for all agents not as some truth being somehow imposed on the will from outside, but as a necessary part of the solution to a practical problem that stems from the will itself, which we as agents must deal with and solve. To repeat: "If you recognize the problem to be real, to be yours, to be one you have to solve, and the solution to be the only or the best one, then the solution is binding upon you."[36]

V. CRITIQUE OF THE RECONSTRUCTED KANTIAN ARGUMENT

Does the argument succeed in delivering such results? Premises 1 and 2 are plausible enough: even Korsgaard's realist opponents should be happy to grant that the practical problem she has identified is a real one that imposes on us the requirements described in those two premises. And so far the sense in which we must do what is claimed is clear: it is a matter of literal practical necessity in the sense that one literally cannot function as an agent without acting on reasons, and one cannot act on reasons without regarding some considerations as good reasons for acting in certain ways (given what it is to act on reasons), and one cannot do that without regarding some ends as worth pursuing (given what it is to judge a consideration to be a good reason for acting). At this point, however, the continuation of the argument may be interpreted in two very different ways, owing to an ambiguity in the claims of necessity.

On the first interpretation, the argument continues to employ this

35. See Korsgaard, "Kant's Formula of Humanity," 123: "If you view yourself as having a value-conferring status in virtue of your power of rational choice, you must view anyone who has the power of rational choice as having, in virtue of that power, a value-conferring status"—and so similarly with viewing oneself or one's humanity as valuable: "Valuing humanity in your own person rationally requires valuing it in the persons of others" (Korsgaard, *The Sources of Normativity*, 121). Korsgaard actually puts the latter claim conditionally because of a potential objection she goes on to consider and answer in the fourth lecture. I shall ignore those complications here, as I am willing to grant the claim in question.

36. Korsgaard, "Realism and Constructivism," 116. Compare Korsgaard, *The Sources of Normativity*, 114.

same notion of literal practical necessity in premises 3 and 4, with the aim ultimately of securing the normative force of the formula of humanity by showing that it is likewise literally practically necessary for us—an enabling principle or constitutive norm without which we simply cannot function as agents and would lose our very existence or identity as unified agents or selves.[37] Much of Korsgaard's recent discussion of normativity in terms of literal practical necessity—for example, the claim that "the necessity of conforming to the principles of practical reason comes down to . . . the necessity of acting," which is "the simple inexorable fact of the human condition"—would seem to require an interpretation along these lines.[38] The constructivist derivation of the formula of humanity is, after all, supposed to show precisely why this principle has normative force for all agents as a principle of practical reason (rather than merely deriving it as some truth whose normative force remains to be seen). But Korsgaard conceives of "principles of practical reason" as "constitutive standards of actions," where this means "standards that we must be at least trying to follow if we are to count as acting at all," thus "explain[ing] their normativity."[39] Thus, the constructivist derivation of the formula of humanity would have to show this principle to be normative for us by showing that it is something we must at least be trying to follow if we are to succeed in exercising agency at all. But the only way the argument could show that would be by showing, through a series of claims about what is literally practically necessary for successive conditions of agency, that the formula of humanity is a principle we literally must at least try to follow if we are to succeed in exercising agency at all.

So interpreted, the argument would fit neatly with Korsgaard's emphasis on the literal practical necessity of the principles of practical reason for agency and with her practical-problem-solving conception of normative force, since our employing the formula of humanity would have been shown to be a necessary solution to our practical problem of agency. Moreover, her most recent remarks specifically about the formula of humanity clearly align with such a focus on literal practical necessity: respect for one's humanity is a "condition of action" without which "it is impossible to will the laws of your own causality, to make something of yourself, to be a person; and unless you make something of yourself, unless you constitute yourself as a person, it will be impos-

37. Korsgaard, "The Normativity of Instrumental Reason," 247–50, and "Self-Constitution," lecture 1, 17–19, and lecture 2, 1. See also Korsgaard, *The Sources of Normativity*, 236: "My own aim is to portray moral principles as constitutive of, and so as essential to, making human choices, and leading a human life."

38. Korsgaard, "Self-Constitution," lecture 1, 17, 2.

39. Ibid., lecture 2, 1. See also n. 37 above.

sible for you to act"; indeed, "respect for the humanity in your own person, and that of every other, is a necessary condition of effective action"—where "effective action" means genuine action that can be attributed to you, as opposed to mere movements attributable only to "forces working in you or on you."[40]

The difficulties with the constructivist derivation as understood on this first interpretation, however, are so extensive (as brought out in Sec. V.*A*) that it is imperative to consider whether there is another, more plausible interpretation of it. And in fact there is. The second interpretation, which more likely reflects Korsgaard's original understanding of the argument considered in itself, sees it as involving a shift in the sense of the necessity claims after the first two premises: while the 'must' in the first two premises is indeed that of literal practical necessity for agency, it thereafter shifts to something weaker, having to do with the idea of commitment, leaving us with a correspondingly weaker conclusion. So understood, the argument does not establish the literal practical necessity of accepting or following the principle that is ultimately derived but shows only that we are *committed* to the principle by certain other things that are literally practically necessary for agency, as described in the first two premises. This interpretation of the argument will be considered in Section V.*B*.

I shall argue that on either interpretation the argument has serious problems in its own right. In the case of the first interpretation, which seemed necessary in order to mesh with the recent emphasis on the practical necessity of normative principles, these problems seem insurmountable. If so, then there is an incoherence between much of Korsgaard's general account of normativity in terms of practical necessity for agency and the constructivist account of the fundamental principle of morality. The second interpretation offers a better construal of the argument itself, but the problem then is that even if this second version could overcome the other difficulties I will raise, it would fail to satisfy Korsgaard's own criteria for binding normative force in terms of the practical-problem-solving model, at least in the manner she claims has marked advantages over realism. We will return in Section VII to consider another way this model of normativity may be applied to such an argument, which I believe is the most promising direction for Korsgaard's positive project. In the end, however, I shall argue that at best this would leave her no better off than the realist, undermining her claim of superiority for constructivism in accounting for normativity.

40. Korsgaard, "Self-Constitution," lecture 6: "Integrity and Interaction," 25–26.

A. Problems for the First Interpretation

If we were to take the argument as employing a consistent sense of practical necessity, then we would have to understand premises 3 and 4 as claiming that just as an agent literally cannot exercise her agency unless she regards or sees some reasons and ends as good (premises 1 and 2), so too she literally cannot regard or see any reasons and ends as good—and so again cannot solve her practical problem of agency—unless she regards or sees herself as the unconditionally valuable, value-conferring source of the value of her ends. That is, she must, as a matter of literal practical necessity, so regard herself.[41]

It is not hard to see, however, that nothing Korsgaard says provides support for such a claim, which is untenable in any case. On this reading, the claim in question is a psychological claim about how an agent must regard things as standing with respect to value and its source in order to exercise agency—just as the first two premises are psychological claims about how we must regard certain things as standing with respect to reasons and ends. This psychological claim does not follow from any value-theoretic claims about how things in fact stand with respect to the source of value. In explicating this argument, however, it is the latter that Korsgaard cites for support, claiming that Kant began with "the relativism of value to human desires and interests" via our choices: "[Kant] asked what it is that makes these objects [of our choices] good, and, rejecting one form of realism, he decided that the goodness was not in the objects themselves" but has its source in our choices.[42] Even if this contentious antirealist claim about the actual source of goodness turned out to be true, its truth would not imply anything about the psychological necessity for an agent to believe it, on a par with the earlier psychological necessity for an agent to regard some ends as good.

Nor do we find support for the relevant psychological claim in the thought that "were it not for our desires and inclinations . . . we would not find their objects good" or that "we take things to be important

41. Korsgaard herself emphasizes that while she does believe (and takes Kant to believe) that all value is conferred—"the result of our own acts of conferring value"—she does not take this to be some kind of metaphysical fact, "as if value were some sort of metaphysical substance that got transferred from us to our ends *via* the act of choice," but takes it instead to be a point about the nature of valuing ("Motivation, Metaphysics," 63–64). The claim is that it is in some sense a condition of our valuing our ends that we "see ourselves as value-conferring," and this in turn requires that we "place value upon ourselves" (60–61, 57).

42. Korsgaard, *The Sources of Normativity*, 122, and (on reflective awareness, choice, and inclination) "Motivation, Metaphysics," 51. See also Korsgaard, "Kant's Formula of Humanity," 121–22.

because they are important to us."[43] (The latter is ambiguous between 'we take it to be the case that things are important because they are important to us' and 'it is insofar as things are important to us that they strike us as being important'. Since the first is obviously false as a generalization about what all of us take to be the case about value, we may assume the second is intended.) Others have already pointed out that from the uncontroversial fact that "we would not find" things to be valuable if we didn't have certain relevant desires and interests, nothing follows about values being dependent on these desires and interests, either directly or via exercises of choice based on such inclinations.[44] But the present point is that whatever the ultimate truth about the source of value, nothing follows about the psychological necessity of *regarding* or *seeing* ourselves as value-conferring sources of value. On the present interpretation, however, Korsgaard would be making just such a mistake. For she maintains that it is from precisely the thoughts quoted above that Kant "concluded that we must therefore take ourselves to be important" and that "humanity, as the source of all reasons and values, must be valued for its own sake";[45] thus, if these conclusions are understood as claims about what is psychologically necessary for us in connection with the exercise of agency, as they are under the first interpretation, then Korsgaard would be making the above erroneous inference.

   Not only are premises 3 and 4, so understood, not supported by any of the considerations Korsgaard cites but they are untenable in their own right. Korsgaard claims that, through Kant's argument, a person can be "moved to the realization that he regards himself as the source of value."[46] But while it may be true of some agents that they regard themselves as the source of value, it is manifestly false of many of us if understood as a psychological claim about how we regard or see ourselves or how we must regard ourselves in order to exercise agency. Rightly or wrongly, many people think that fidelity or piety, for example,

43. Oddly, she also says at one point that "Kant urges us to take things to be important *because they are important to us*" (Korsgaard, *The Sources of Normativity*, 242). Saying that Kant "urges" us to do this is obviously different from claiming that "Kant saw" that we do this (*The Sources of Normativity*, 122).

44. See, e.g., Allan Gibbard, "Morality as Consistency in Living: Korsgaard's Kantian Lectures," *Ethics* 110 (1999): 140–64, esp. 147–50. Gibbard points out that Kant's (alleged) reflective subjectivist conclusion about the source of value—e.g., that "biodiversity wouldn't be important unless we reflectively desired it"—does not follow from the claim that "we wouldn't think biodiversity important unless we reflectively desired it." Indeed, realists— and even expressivists—can certainly grant that we would not find the objects of our desires good if we didn't have those desires, but nothing follows about those objects' thereby not being good.

45. Korsgaard, *The Sources of Normativity*, 122.

46. Korsgaard, "Motivation, Metaphysics," 62.

are important simply because God has decreed them so or that humane treatment of animals is important simply because their suffering is intrinsically bad. Even if these beliefs turn out to be false, it is incorrect to say straightforwardly of such people that they "regard [themselves] as the source of [the] value" of fidelity or kindness, or "see [themselves] as value-conferring" in relation to these ends; and it would be false to say that they must do these things in order to be able to view fidelity or kindness as ends worth promoting and thus solve their practical problem.[47] Their views—whether true or false as philosophical accounts of the source of value—suffice to solve their practical problem, as set forth in the first two premises. All they need for the purpose of solving the generic practical problem of agency is *some* way of regarding some ends as worth pursuing, and thus some considerations as good reasons for acting, so that they can act on what they consider to be good reasons and thus exercise agency.

There is, then, no purely practical need for agents to adopt the stance Korsgaard describes in steps 3 and 4, in the sense that doing so would be literally practically necessary for the exercise of agency. On this first interpretation, the argument thus stalls out immediately after the first two premises and has certainly not shown that agents must, as a matter of literal practical necessity for agency, follow or try to follow the formula of humanity. Perhaps this need not really be the goal of the constructivist derivation of the formula of humanity. If it is not, however, then Korsgaard will have to drop her earlier quoted claims of practical necessity in connection with the principles of practical reason, at least when it comes to the formula of humanity. We will consider later the extent to which this concession would damage her project.

B. Problems for the Second Interpretation

Upon reflection, the first interpretation of the argument had little plausibility, despite its naturally fitting much of Korsgaard's language and her general claims about the principles of practical reason. There is, however, a more promising interpretation of the argument. On this second interpretation, the point is that an agent—whatever he believes explicitly about the source of the value of his ends—must in any case *treat himself as if* he were it.[48] That is, in pursuing his ends as objects worthy of pursuit, which he does in acting at all, an agent is necessarily

47. Ibid., 60–62, and *The Sources of Normativity*, 122. Guyer notes that "the idea that rational choice itself is value-conferring" is a "philosopher's notion par excellence" rather than something "readily available to every human being" ("The Value of Reason," 28). This supports my rejection of the claim that agents need to employ such ideas in the way they "view" or "see" things in order to be able to exercise their agency at all.

48. Korsgaard uses this sort of construction in "Kant's Formula of Humanity," 123: "We act as if our own choice were the sufficient condition of the goodness of its object."

treating himself as if he were the unconditionally valuable, value-conferring source of the value of those ends, whether he explicitly realizes this or not.[49] But that in turn means that he is practically committed to anything implied by his being such a source of value—in the same way that an agent who must treat herself as if she were free is practically committed to whatever laws would apply to a free will, whatever her theoretical beliefs about freedom. The rest of the argument, then, simply brings out these practical commitments. Once we get past the second premise, the sense of 'must' in connection with claims about how we must regard or see ourselves or others (in 3b, 4b/5, and 6/7 below) shifts to, roughly, 'ought to, in the name of consistency with the practical commitments we take on in acting at all'.

The argument, as it proceeds after premises 1 and 2 (which remain claims of literal practical necessity for agency), may thus be spelled out as follows:

3a.   In regarding some ends as being genuinely good, we are necessarily treating ourselves as if we [qua possessors of rational nature] were the value-conferring source of the value of those ends, thereby practically committing ourselves to the latter and to its implications.

3b.   Thus, we *must*—that is, ought to, in the name of consistency with our practical commitments—"regard ourselves as capable [by virtue of our rational nature] of conferring value upon the objects of our choice, the ends that we set," that is, we ought to "see ourselves [qua possessors of rational nature] as value-conferring" and "regard [*ourselves*] as the source of [the] value" of those ends.

4a.   In so regarding ourselves, we would necessarily be treating ourselves as if we were important—indeed unconditionally valuable qua possessors of rational nature—thereby practically committing ourselves to this and to its implications.

4b/5.   Thus, we *must*—that is, ought to, in the name of consistency with our practical commitments—regard ourselves as unconditionally valuable qua possessors of rational nature.

6/7.   Therefore, we *must*—that is, ought to, in the name of consistency with our practical commitments—regard other beings equally possessed of rational nature as likewise being unconditionally valuable ends in themselves.

49. See Korsgaard, *The Sources of Normativity*, 249: "Guided by reflection, we may be led to see that our tendency to regard our desires as reasons implies that we set a value on our own humanity." According to the present interpretation, "setting a value on our humanity" here means that we are treating our humanity as if it were the valuable source of our reasons, whether or not we explicitly so value it (though we would if we reflected sufficiently).

Considered in itself, this version of the argument seems more promising. Whether it truly fares better, however, will depend on whether it has really been shown that, in the course of solving the practical problem in the first two premises, one would necessarily be treating oneself as if one were the unconditionally valuable, value-conferring source of the value of one's ends (3a and 4a). To answer this question, we need first to clarify the required notion of treating something as if it had a certain property by acting in a certain way. The general idea seems to be this: in acting in a certain way, one treats X as if it were F insofar as the action is predicated on one's taking it to be the case that p, and p could in fact be the case only if X were F. So insofar as any action of mine is predicated on my taking it to be the case that my end is good, and my end could be good only if I were an unconditionally valuable source of its value, I am—simply by so acting—treating myself as if I were an unconditionally valuable source of the value of my end. The burden of the argument would thus be to show that it is indeed the case that

S: My ends could be good only if I were the unconditionally valuable source of their value.

Only if this is so can it be said that in pursuing ends I judge to be good I am treating myself as if I were the unconditionally valuable, value-conferring source of the value of my ends.

Now S is a value-theoretic claim about the true dependency of the value of ends on agents. Ironically, this sounds like just the sort of metaphysical construal of value conferral that Korsgaard herself renounced in favor of claims simply about how we allegedly must see and value things and ourselves.[50] The latter, of course, was the focus of the first interpretation of the argument, which failed. We are thus back to what looks like a crucial metaphysical claim about how things truly stand with respect to the source of value.

Setting this tension to one side, we may notice that, in relying on S, the positive constructivist argument depends entirely on the success of sweeping negative arguments against realism. For if even a very modest form of realism were true, S would be false. If, for example, severe animal suffering is intrinsically bad—bad in a way that is not derived from facts about the conditions of my exercise of agency—then my end of stopping a forest fire could be good quite apart from any value conferral on my part. Unless we have been given independent reason to reject such claims, we have no reason to accept S. It is not obvious that Korsgaard's general attack on realism (Sec. II) will help here, since that was focused on the problem of normative force rather than on the present question whether ends can be good in a relevantly nonderivative

50. Korsgaard, "Motivation, Metaphysics," 63–64. (See n. 41 above.)

way, violating S. And in any case, one of my aims is precisely to deflate her critique of realism. If that is successful, then, it will not only answer her objections to realism but will also undermine support for the positive constructivist argument as presently construed, insofar as that argument depends on the prior rejection of realism.

Apart from relying on an independent rejection of realism, Korsgaard's support for S seems to come down to a simple intuition: if I didn't matter, then it couldn't matter whether my ends were realized or frustrated; if it matters that my ends be realized, then that implies that I am an important being.[51] This reasonable intuition, however, has force only in connection with ends having to do with my own welfare and fulfillment. My end of taking piano lessons in order to develop musically, for example, would indeed not matter if I were worthless and my welfare counted for nothing. But there is no similar intuitive pull to say anything parallel about any number of other ends I might pursue, as in the forest fire example. And even with respect to ends that do pertain to the agent's own welfare, all that is compelling is that the agent-neutral importance of the ends' being realized is conditional on the importance of the agent or of the agent's welfare. Nothing follows about the agent's being the value-conferring source of the choice worthiness of the ends themselves in the sense that what makes these ends worth pursuing is explained by the agent's desires, inclinations, or choices together with the fact that the agent is unconditionally valuable and endorses those desires or makes those choices.[52]

Let us, however, again set these substantial problems to one side. The more fundamental and important problem is that the argument, as presently construed, has apparently abandoned the larger model of normativity within which it was supposed to be working (though see Sec. VII for an attempt to remedy this problem). Korsgaard's model of normativity is clearly employed in the first two premises, which describe a practical problem and what we must do to solve it, but after that the argument shifts markedly, and what we get is something of a different

51. Ibid., 62: "Imagine it this way: Why in the end does it matter that I achieve this end? Because it matters to me and I matter."

52. Korsgaard, *The Sources of Normativity*, 122, and "Kant's Formula of Humanity," 121–24. In "The Reasons We Can Share" (reprinted in her *Creating the Kingdom of Ends*, 275–310), Korsgaard offers further support for such an intersubjectivist position, setting it against what she calls "objective realism," which is an extreme, Platonic realism. While she succeeds both in raising problems for the latter (with its "intrinsic values already there in the universe," 290) and in making plausible the intersubjectivist model for the value of chocolate or of mountain climbing, she provides no compelling argument that the intersubjectivist model is appropriate for all cases. An alternative would be a pluralistic model, with intersubjectivism applied to a certain range of cases, and a moderate realism (perhaps along neo-Aristotelian lines) applied to others.

nature: not that we must accept or conform to principles concerning the value of humanity in order to solve our practical problem and exercise agency, but rather that our exercise of agency commits us to those principles, such that if we fail to accept or to conform to them we can be criticized as being *inconsistent*. In acting at all, we allegedly cannot help committing ourselves to certain principles and thus to the duties implied by them. So the real upshot is this: if we are to act in accordance with the true implications of our exercise of agency, according to the correct philosophical account of such things, then we must accept and conform to the formula of humanity on pain of inconsistency if we do not.[53]

Thus, even if S were given more compelling support than Korsgaard has offered, the most that will have been shown is that we are *committed* to accepting the formula of humanity, or to conforming to the moral requirements derived from it, and so "must" do so in this sense. This is roughly the same sense in which someone who believes 'p-or-not-q and not-p' is committed to believing (or "must" believe) 'not-q': he is guilty of a certain inconsistency if he does not. There is a difference in that the commitment Korsgaard focuses on is rooted in practical agency rather than in beliefs, and it is a commitment to something practical. But none of this makes the necessity in question practical in the relevant sense with which we began, that is, a matter of what we must do in order to be able to exercise agency at all. Such an argument therefore does not support the claim that the formula of humanity is an enabling principle we must at least try to follow in order to solve our practical problem and count as acting at all.

There is, then, a substantial disconnect between the specific argument for the normativity of the fundamental principle of ethics, on the one hand (simply in terms of commitment), and Korsgaard's general insistence, on the other hand, that normativity is to be understood as a form of necessity that "comes down to the necessity of being an agent . . . the necessity of acting," which is a literal practical necessity that she takes to guarantee the binding normative relevance of a principle to an agent's will in a way that realism cannot.[54] The claim that accepting

53. This shift away from claims of literal practical necessity after the beginning was in fact clearly present all along in premise 6: for the 'must' in premise 6 is clearly about what is required if one is to be fully consistent, rather than about what is strictly practically required in order to be able to exercise agency at all, as in the first two premises. Indeed, this is clear in one of Korsgaard's formulations of premise 6, where she says that "valuing humanity in your own person rationally requires valuing it in the persons of others" (Korsgaard, *The Sources of Normativity*, 121). This further undermines the argument under the first interpretation in Sec. V.*A*, showing that it would have failed to go through as a chain of literal practical necessities even if it had not already floundered at premise 3.

54. Korsgaard, "Self-Constitution," lecture 1, 17.

and conforming to the formula of humanity is necessary if one is to avoid inconsistency with what philosophical investigation reveals to be the implications of one's exercise of agency is a far cry from the claim that accepting and conforming to the formula of humanity is practically necessary in order to solve our original practical problem of agency so that we may act at all.

C. Shifting Necessity Claims and the Normative Question

This raises a significant difficulty in connection with the larger issue described in the first three sections. The problem is that the shift to the weaker conclusion about commitment opens up precisely the sort of normative gap Korsgaard was so concerned to close with her practical-problem-solving approach in the first place. For we may now ask: why should I care about acting fully consistently with all the implications of my exercise of agency? And why should I care about this more than I care about anything else, especially when some onerous sacrifice may be involved? To put the point in the same terms in which Korsgaard framed her critique of realism: if it is a *fact* that I will be acting consistently with all the implications of my exercise of agency only if I make some sacrifice required by the formula of humanity, then it is so far an open question whether or not I should apply that fact in my deliberation and give it overriding weight. Why should I? Her own argument against realism may thus be turned around against her own account, again raising the prospect of an infinite regress of justifying norms.

Korsgaard would no doubt like to be able to respond by claiming that we as agents should care about these issues of commitment and consistency because we have no choice: they are practically necessary for our very functioning as agents, which is our "plight" as creatures with reflective consciousness. This would provide a final answer to the question about why we should care. But while such claims of necessity for agency are plausible in connection with the issues in the first two premises—that is, that we must regard some reasons as good and some ends as worthy of pursuit—it is a different question whether similar claims of literal practical necessity can be made out in connection with the issues of commitment and consistency. What Korsgaard would have to be able to show here is that it is literally practically necessary in order for us to exercise agency at all (1) that we strive to act fully consistently with all the implications of the exercise of our agency (as revealed by Korsgaard's and Kant's sophisticated philosophical reflection), and it is similarly necessary (2) that we care about this more than we care about all the things that might conflict with those implications, such as a central life project the pursuit of which might happen to conflict with the formula of humanity.

This cannot be shown, however, because again it is false. The prac-

tical problem of agency at the heart of Korsgaard's project, articulated in the first two premises, is one that is solved every day as a practical problem by people who fail to act consistently with all the commitments Korsgaard would claim are actually implied by any exercise of agency. Many people, for example, exercise their agency for whole lifetimes while regularly failing to "share others' ends" in the robust way indicated in the formula of humanity as Korsgaard understands it. This alleged failure of consistency plainly does not undermine agency in anything like the way that a thoroughgoing nihilism might. Inconsistency may give our actions the evaluative quality of being bad, but that is nothing more than the realist would say and doesn't secure the kind of agency-based practical necessity that Korsgaard seems to be after in order to set herself apart from the realist and from his reliance on irreducibly evaluative claims.

Korsgaard does make a plausible case for the practical necessity of avoiding too much inconsistency of certain kinds, arguing, for example, that a general failure to will the means to one's willed ends or to adhere to any coherent conception of practical identity will result in a disintegration of agency.[55] Such arguments, however, address only the normativity of the statistical claims, such as the claim that one ought to avoid too many failures to take the means to one's ends. They cannot address the normativity of the instrumental principle itself, for example, which tells us for *each* of our ends that we ought to avoid failing to take the known, necessary, available means while continuing to will the end. The instrumental principle may well be understood as rooted in a point about consistency, but the attempt to underwrite this with literal practical necessity for agency (through statistical claims about what is needed for the maintenance of unity) does not amount to accounting for instrumental normativity in the particular case in terms of practical necessity.[56]

Returning to the formula of humanity, nothing in the Kantian argument so far supports even the statistical necessity claim analogous to the claim about the need to avoid too much instrumental irrationality. If the Kantian claims about what we commit ourselves to in any exercise of agency turn out to be correct, then people who fail to act consistently with those commitments will thereby be defective. But this defect cannot be reductively described in terms of their having failed to solve, for

55. See Korsgaard, "The Normativity of Instrumental Reason," 247, and *The Sources of Normativity*, 101–2, respectively. Compare also much of Korsgaard, "Self-Constitution"— e.g., lecture 2, 17, where an agent's obligation to conduct himself in accordance with a certain principle is identified with the fact that if he does not, then "he will not have a mind at all."

56. I discuss this in detail in my "Normative Realism, Kantian Constructivism, and the Instrumental Principle" (unpublished manuscript).

practical purposes, the generic practical problem with which we started.[57] Such agents act badly, but they act, and they needn't and typically don't disintegrate as agents. This might happen in extreme cases such as that of Plato's tyrant, but it is a familiar objection to Plato's argument that one cannot show that a moderately unjust life is psychically disastrous simply by showing that a maximally unjust life is. Undermining the tyrant's claim to happiness does not show that each of us is always better off acting justly and so does not explain the normative force of particular requirements of justice. Similarly, even if the Kantian argument did somehow show that a sweeping rejection of the formula of humanity would make it impossible for an agent to solve her practical problem of agency, this would not succeed in accounting for the normative force of moral requirements in particular cases.

Korsgaard might respond to this line of argument by weakening her position to the following: people who fail to act consistently with all the commitments arising from their exercise of agency are insufficiently reflective, and if an agent were sufficiently reflective, he would grasp all such implications and not attempt to deceive himself about them, and in that case it would be necessary for him to act consistently with all of these commitments if he is to maintain his very integrity or coherence as a self-conscious agent. Thus, if someone does not act consistently with such implications, then he can maintain his agency only at the cost of being "condemned as insufficiently reflective."[58]

This, however, still fails to get around the problem: for the question now is why an agent must care above all else about being "sufficiently reflective" in the relevant sense. He must, of course, be reflective enough to exercise agency if he is to act at all, but that doesn't imply being "sufficiently reflective" in the relevant sense that would entail grasping all of the alleged implications of exercising agency (as even Korsgaard's quite reflective philosophical opponents have evidently failed to do). Nor must an agent be "sufficiently reflective" in order to keep his agency intact, since the above concern over integrity of agency arose only given that the agent is sufficiently reflective: if he is not, then there needn't be any cognitive dissonance resulting from his failure to do everything he is committed to. So it has not been shown that an agent must care— and care above all else—about being sufficiently reflective. And Korsgaard cannot get around this difficulty just by restricting the account

57. For such reductive claims, see Korsgaard's remarks that "what makes actions good or bad is how well they constitute you" and "a good person is someone who is good at being a person," where this means simply: successful at maintaining one's unity and identity as an agent, i.e., solving the practical problem of agency ("Self-Constitution," lecture 1, 17–18).

58. Korsgaard, *The Sources of Normativity*, 250.

of normativity to people who do care about being sufficiently reflective and so care about living consistently with all their agency-related commitments. For her whole project is to account for the nonhypothetical normative force of moral requirements—the categorical bindingness of the formula of humanity on all agents. To restrict the scope of the argument so that the formula of humanity has normative force only for agents who happen to care about being sufficiently reflective in her sense would be to abandon the essence of the project, which is Kantian not only in its focus on agency but also in its commitment to the categoricity of moral requirements. Indeed, as we have seen, Korsgaard elucidates "the normative question" precisely by imagining people who are powerfully tempted to subordinate morality to other values when moral requirements become onerous. Her question is how moral requirements have normative force for agents even in such situations, and it would altogether miss the point to attempt to answer that question with an argument that relied on the assumption that the agent happens to care above all else about acting fully consistently with all of the commitments stemming from the exercise of his agency.

It is likewise not an option for Korsgaard simply to label people as "irrational" if they do not care above all else about being fully consistent in this way and to say that her account of normativity need only speak to agents insofar as they are fully "rational." Such a "stipulative" solution—analogous to the realist's labeling people as "irrational" if they don't care sufficiently about performing actions that can be justified as being good—is no solution at all to the kind of problem Korsgaard herself has pressed, and she recognizes this.[59] Indeed, if it were that easy, the entire appeal to practical necessity for agency that dominates her recent work would be superfluous: the job would be complete with the familiar points about consistency. For similar reasons, Korsgaard cannot simply claim that facts about consistency or fully reflective agency are intrinsically normative. Such a claim is not implausible, but it would be in obvious tension with her earlier critique of realism, where it is assumed that it makes perfect sense for an agent to ask such questions as why she should care about the goodness or badness of her action and that we need to have a persuasive answer to such questions if we are to account for normativity. If such questions about goodness are supposed to be legitimate and pressing, then it is surely fair to press the parallel question why an agent should care so much about being fully reflective and consistent. And telling a hard-pressed agent that facts about consistency are intrinsically normative is no more help coming from Korsgaard than is the parallel claim about facts concerning goodness of action coming from the realist. Facts about commitment

59. See, e.g., Korsgaard, "The Normativity of Instrumental Reason," 219 n. 11.

and consistency may well be intrinsically normative, but Korsgaard cannot help herself to such a claim without forfeiting her advantage over the realist and his intrinsically normative facts.[60]

There would, of course, remain one important difference: by starting with the practical problem and its solution in premises 1 and 2, she would be anchoring her further appeals to facts about commitments, tying them to the agent's will in a practical way rather than simply relying on them by themselves in the way that realists have traditionally appealed simply to intrinsically normative facts about goodness. This is an important point, which I will develop in Section VII. The apparent advantage, however, is again illusory. As we shall see in Sections VIII and IX, a parallel anchoring move is equally open to the realist.

VI. KORSGAARD'S "FANCY NEW MODEL"

Having examined Korsgaard's reconstruction of Kant's argument and evaluated it in light of broader issues concerning normativity, we may deal more briefly with Korsgaard's own, more complex version, for which very similar difficulties arise. The argument again begins with the agent's need for reasons, that is, for some normative conception of what there is good reason to do, and proceeds in the same constructivist spirit, using very similar language. What is new is just that the solution to the initial practical problem is now taken to revolve around the idea of an agent's "conception of practical identity," since "a view of what you ought to do is a view of who you are."[61] The argument may be laid out as follows:[62]

1. Any human being, because of the reflective nature of her consciousness, requires reasons for acting and hence *must* have some normative conception of what there is reason to do in order to be able to act at all.
2. To act from such a conception just is to act from a conception of

---

60. It might be thought that something like consistency is a more plausible candidate for an intrinsically normative concern because it is wholly internal to the agent and so of more obvious concern to her. But the goodness to which the realist appeals is also internal to the agent in a relevant sense, since we are talking about the goodness of action as performed by the agent—the goodness of the will's own activity. The fact that this goodness is construed in a realist, rather than constructivist, fashion does not diminish this intimate connection to the agent, and the externality implied by realism should therefore not be exaggerated—a point I develop in William J. FitzPatrick, "Reasons, Value and Particular Agents: Normative Relevance without Motivational Internalism," *Mind* 113 (2004): 285–318.

61. Korsgaard, *The Sources of Normativity*, 117. See also 100–101.

62. The argument I have here reconstructed is given primarily in Korsgaard, *The Sources of Normativity*, 113–30 (esp. 120–23), 250. Quotations are from 123 unless otherwise noted.

one's practical identity, which one takes to be normative—"a conception under which you . . . find your life to be worth living and your actions to be worth undertaking."

3. Thus, any human being *must* have some particular conception of her practical identity that she takes to be normative and allows to govern her action—a requirement having its source simply in her humanity, that is, coming "from [her] identity simply as a human being, a reflective animal who needs reasons to act and to live."[63]

4. But if an agent is to take her particular conception of her practical identity to be normative (i.e., as giving her genuine reason to conform to it), she *must* also take her humanity—that is, her identity as a reflective animal who needs reasons to act, which is the source of her requirement to have some more particular conception of her practical identity—to be a normative form of identity and thus value it.[64]

5. Thus, an agent "*must* take this . . . identity [i.e., her humanity] . . . to be normative" and so "*must* value [her] own humanity if [she] is to value anything at all" (which latter she in turn must do, again, in order to act at all).[65]

6. Therefore, an agent *must* likewise value the same humanity in others, taking it to be normative, as articulated by the formula of humanity.[66]

As with the earlier argument, this one begins with a practical problem of agency, laid out in steps 1–3, which employ a clear notion of literal practical necessity. Just as an agent must regard some ends as worth pursuing if she is to act on reasons and exercise agency, so too she must have some conception of her practical identity that she takes to be normative. A realist can agree that, without such a normative conception, an agent will "lose [her] grip on [herself] as having any reason to do one thing rather than another . . . [and so] as having any reason to live and act at all."[67] The difficulties begin with premise 4.

What makes a particular conception of practical identity normative for an agent, giving her genuine reasons to do certain things? Kors-

---

63. Ibid., 121. Compare 123: "Since you are human you must take something to be normative, that is, some conception of practical identity must be normative for you. If you had no normative conception of your identity, you could have no reasons for action, and because your consciousness is reflective, you could then not act at all." One thus has "no option" but to accept this requirement.

64. Ibid., 121.

65. Ibid., 123 (emphasis added). Similarly: "Since you cannot act without reasons and your humanity is the source of your reasons, you must value your own humanity if you are to act at all" (123).

66. Ibid., 121.

67. Ibid.

gaard's answer is the fact that this conception of identity is what is fulfilling the generic and inescapable requirement imposed by her humanity to have some such conception of identity. This is why the agent has reason to take seriously this conception of identity and the reasons bound up with it (unless she is prepared to replace it with some other conception of her identity, in which case the same point would then apply to that one): she must conform to it "because being human requires it."[68] The agent's humanity is thus "the *source* of [her] reasons," since it is the source of the necessity for her to take the reasons offered by her particular identity seriously—a necessity Korsgaard thus identifies with normative force, making those purported reasons genuine normative reasons for the agent.[69] And if an agent's humanity is the source of her reasons, then it is a normative form of identity and has value. How, after all, could something serve as the source of reasons if it did not itself possess normative authority and therefore value?[70]

Even if this were correct, however, so that "all value [thus] depends on the value of humanity" as Korsgaard claims,[71] the fundamental question is how this would support premise 4 and hence the subconclusion in 5. As with Kant's argument, there are two ways one might proceed here.

The first parallels the first reading of Kant's argument (Sec. V.*A*), taking the claims in steps 4 and 5 as strict necessity claims about the agent's psychology—claims about what she literally must take to be normative, or must value. This is not an unnatural reading, inasmuch as the same language of necessity (i.e., what any agent must do) is used in just this sense in the first half of the argument, with no acknowledgment of any shift in the sense of the modal claims in the second half. But as we saw in the case of Kant's argument, if this is how the argument is to be understood, then steps 4 and 5 are simply unsupported. Even if an agent's humanity were indeed the true source of his reasons in Korsgaard's sense, he needn't take this to be so in order to take his

68. Ibid.

69. Ibid., 123.

70. This is a stronger and more charitable interpretation of Korsgaard's argument than Gibbard's, e.g.: "The premise, 'Value being a brave Achaean!' has the form, 'Value X!' Next we have 'X entails Y,' being a brave Achaean entails being human. The conclusion is 'Value Y!' The validity of Korsgaard's argument depends, it seems, on whether this general form fits and is valid" (Gibbard, "Morality as Consistency in Living," 154). I do not think that this is the general form of the argument, which is a good thing since such an argument would also show that we are committed to valuing such things as having mass (as being a brave Achaean—or anything else available to humans—entails having mass). The argument as I have presented it is much richer and more obviously connected with Korsgaard's interpretation of Kant's argument, which illustrates why it was important to look closely at the latter.

71. Korsgaard, *The Sources of Normativity*, 121.

particular conception of practical identity to be normative and thus solve the practical problem laid out in premises 1–3. Perhaps he has been convinced by reading Clarke that what makes his conception of practical identity normative and thus reason providing is the fact that it conforms to the "eternal fitnesses," which is why he adopted it in the first place. Even if this turns out to be a false philosophical doctrine, it suffices as far as the purely practical issue is concerned, and he is able to exercise his agency by acting on considerations he takes to be reasons. So it has not been shown that an agent must take the normativity of his conception of identity to be rooted in the mere fact that this identity is fulfilling for him the need imposed by his humanity for him to have some conception of identity. The argument again stalls out.[72]

Consider, then, the alternative reading, paralleling the second interpretation of Kant's argument (Sec. V.*B*), according to which the sense of the necessity claims shifts halfway through. On this reading, the argument proceeds as follows after step 3:

4′. In taking her particular conception of practical identity to be normative (i.e., as giving her genuine reason to conform to it), an agent treats her humanity as if it were a normative form of identity, serving as the ultimate source of her reasons.

5′. Therefore, an agent cannot help treating her humanity as if it were a normative form of identity, serving as the ultimate source of her reasons—which thus commits her to the normativity or value of her humanity.

6′. And since "valuing humanity in [one's] own person rationally requires valuing it in the persons of others," any agent is thus committed to valuing humanity generally and hence to the formula of humanity.[73]

As with the analogous interpretation of the earlier argument, there are two central critical points to press here. First, premise 4′ obviously requires further support in the form of a theoretical claim about how things truly stand with regard to value and its source:

S′:   An agent's particular conception of practical identity could be normative (i.e., giving her genuine reason to conform to it)

72. The argument might work in the special case of someone who happens to share Korsgaard's philosophical views about normativity: if one believes that one's conception of identity could have normative authority for one only insofar as it is filling a need imposed by one's valuable humanity, then perhaps one will have no choice but to so regard and value one's humanity. But as a general argument about how any agent as such must regard and value things, it fails—and it would fail even if Korsgaard's value-theoretic claims about the true source of reasons turned out to be true, just as we saw with the earlier argument.

73. Korsgaard, *The Sources of Normativity*, 121.

> only if her humanity itself is a normative form of identity, serving as the ultimate source of her genuine reasons.

Only if S′ is true can it be claimed that in taking her particular conception of practical identity to be normative an agent is treating her humanity as if it were the normative source of her reasons (see Sec. V.*B*). The real burden of the argument, then, rests on the defense of the value-theoretic claim S′. Not only do we not find real support for this claim, however, but there are good reasons to be doubtful of S′ itself, especially in the present context. Ironically, if we were presupposing a realist framework, it might indeed be plausible that if some entity E is the source of reasons that have genuine normative force, then E must itself have some sort of normative authority and value. But there is little motivation to say the same outside of a realist framework, where all that has been established is that an agent's humanity is the source of her *need* to make normative judgments and so of her need to treat certain things as normative reasons. Why must one's humanity be deeply valuable in order to pull *that* off? And again, recall that Korsgaard herself rejects any metaphysical construal of the sense in which one's humanity "confers value," preferring to stick to claims about how we must value things.[74] But this brings us back to the earlier difficulty, since it has not been shown that an agent strictly must regard her humanity as the source of the normativity of her identity and her reasons. All we now have are claims about commitment, which themselves *depend on*— and so cannot help to support—the claim about value dependence.

The second, deeper problem is equally familiar from the critique of the earlier argument. While the first three premises set up a practical problem of agency and bring out strictly necessary conditions for solving it, what comes after are only weaker claims about what an agent commits herself to in the course of meeting those conditions in the first three premises. Thus, whatever its merits when considered in itself, assuming the above problems could be remedied, the purely constructivist derivation of the formula of humanity fails to accomplish what it is supposed to accomplish according to Korsgaard's own conception of binding normative force: it does not establish the binding force of morality by showing how adopting or conforming to (or even acknowledging) the formula of humanity is practically necessary in order to solve the practical problem of agency with which we began. A conclusion about commitment and consistency is importantly weaker than a conclusion about practical necessity in the sense Korsgaard employs in characterizing normativity. If there is really a normative gap for the realist, then so far it arises for Korsgaard as well.

74. Korsgaard, "Motivation, Metaphysics," 64. See n. 41 above.

VII. PRACTICAL PROBLEM SOLVING WITHOUT PRACTICALLY NECESSARY PRINCIPLES

Korsgaard's central worry about realism was that by merely positing the existence of normative truths confronting an agent—truths external to her own will, imposing themselves on her from outside—we cannot make sense of how such truths get a normative grip on the agent, binding her will; we need some kind of practical connection to anchor normative principles or claims to the will. Her own proposal was to embrace a constructivist account of such principles and claims and to maintain that they are normatively anchored to the will by virtue of being necessary to the solution of an inescapable practical problem of agency: an agent cannot simply shrug them off because they are in some way crucial to solving a practical problem of agency that the agent must solve in order to continue to exist as an agent at all. The obvious question, however, was in what sense the principles are supposed to be necessary for solving the agent's practical problem. Are they necessary in the sense that the agent literally must accept them, or follow them, or both, in order to be able to exercise and maintain her agency—in the same way an agent literally must regard some reasons and ends as good in order to function? Or are they necessary only in the weaker sense that the agent must accept or follow them on pain of inconsistency with commitments taken on in the course of solving the practical problem?

The first of these alternatives fits most naturally with Korsgaard's emphasis on both practical necessity and practical problem solving in her account of normativity. Yet we have seen that the claim in question seems to fail with regard to the fundamental principle of morality itself.[75] The second alternative is more promising as a general strategy, though there remain serious difficulties with the derivation. But the deeper question is how exactly it connects up with the practical-problem-solving model of normativity, given that the principles or claims in question need not strictly be either accepted or conformed to by an agent in order for the agent to solve her practical problem of agency considered simply as a practical problem.

The most promising answer for Korsgaard is to back away from the practical necessity claims, recalling that the key formulation of the prac-

---

75. That is, it has not been shown even that an agent strictly must employ or try to follow the formula of humanity in order to succeed in exercising agency. In "Normative Realism," I argue that Korsgaard comes closer with the instrumental principle: she does plausibly show that an agent who refuses to employ the instrumental principle will not succeed in genuinely willing an end. This, however, falls short of accounting for its normative force in terms of constitutive necessity, since the latter does not sufficiently explain the normative requirement to follow through on the commitment by actually taking the means, which is distinct from the need to employ the principle in willing the end.

tical-problem-solving model is actually disjunctive: "If you recognize the problem to be real, to be yours, to be one you have to solve, and the solution to be the only *or the best* one, then the solution is binding upon you."[76] We have not seen any persuasive argument that accepting and following a principle such as the formula of humanity is the only solution to an agent's practical problem of agency—that is, that it is practically necessary for us. Agents can, and often do, solve their practical problem of agency and avoid "complete normative skepticism" without either accepting or conforming to the formula of humanity, let alone accepting the claims Korsgaard makes about the source of value along the way. But it is still open to Korsgaard to argue that, while accepting or conforming to the formula of humanity is not practically necessary for solving the practical problem of agency, it is nonetheless the best solution to that problem and thus has binding normative force.

She might argue, for example, that it is the best solution because the principle in question is not just any solution to the practical problem: it is one that the agent is committed to by the fact of exercising agency at all. If an agent has a practical problem she must solve and either of two incompatible principles would equally solve it for practical purposes, and she is already committed to one of them, then the best solution would be the principle to which she is already committed rather than the one that would cause her to violate those commitments. Thus, although accepting and following the formula of humanity is not strictly necessary for the agent to solve her practical problem of agency, it is the best solution and so is normatively binding on her. Such a line of argument would preserve the practical-problem-solving approach to normativity while abandoning any attempted reduction of normativity to literal practical necessity for agency. On this approach, it remains an irreducibly normative fact that one principle is a better solution to the practical problem than another and so ought to be followed instead of the other, by virtue of the fact that this is consistent with the agent's commitments—which in turn presupposes the irreducibly normative idea that one ought to act consistently with one's commitments, in this case with all the implications of one's exercise of agency.

Alternatively, Korsgaard might avoid irreducibly normative claims by arguing that what makes one of a number of possible solutions to the practical problem the best solution is that it enables a person to exercise and maintain agency in the most robust way. She maintains, after all, that "some actions are more thoroughly actions than others" and holds that for Kant a morally good agent's "actions are more truly active, more authentically her own, than those of agents who fall short

76. Korsgaard, "Realism and Constructivism," 116 (emphasis added). Compare Korsgaard, *The Sources of Normativity*, 114.

of moral goodness."[77] Even if we grant, however, that some behaviors are more active and spontaneous than others and so are more fully "actions," the burden would be to show that only the favored principles, such as the formula of humanity, provide for the most robust exercise of agency, while rival principles allow us to act and to exist only in a less robust and unified fashion. That would be an interesting result, though it is manifestly not shown by anything in the Kantian derivations of the formula of humanity, which are supposed to establish its normativity for us, nor does it follow from plausible claims (shared by Aristotle) about the degenerate agency that results when we act without reflection or endorsement: people can be morally bad without being passive and unreflective in the face of their impulses. Finally, even if bad agency were ipso facto less robust agency, the question remains why an agent must care in each case about maximally robust exercises of agency more than he cares about anything else he might be called upon to sacrifice—returning us to the original normative question.

There is, however, a more fundamental point to be made about both of the above lines of argument in connection with the debate between constructivists and realists. Even if such moves were successful, they would at most leave constructivism on all fours with realism in accounting for normativity. This is because a directly parallel argument appealing to the "best solution" version of the practical-problem-solving conception of normativity is fully available to the realist. So insofar as the practical-problem-solving model can be made to work at all for Korsgaard, it can be co-opted by the realist precisely to answer her central objection and to provide an account of normativity that is at least as plausible as her own. Let us turn to that.

VIII. REALISM AND NORMATIVITY

What is intriguing about Korsgaard's approach to normativity is the idea of anchoring a normative principle to the will by seeing it as playing a proper role in solving a practical problem for the agent, thus making clear its relevance to the will. Since many principles can evidently serve that practical function, providing an agent with some normative scheme and so allowing her to exercise agency, the natural move to discriminate among such principles was to suggest that a principle has binding normative force if it offers the best solution to the practical problem. Korsgaard's most promising option here is to understand a solution as the best one if it is required in order to follow through on the commitments

77. Korsgaard, "Self-Constitution," lecture 2, 1; Christine Korsgaard, "From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action," in *Aristotle, Kant and the Stoics*, ed. Stephen Engstrom and Jennifer Whiting, 203–36 (Cambridge: Cambridge University Press, 1996), 204.

implied by one's exercise of agency. There is, however, another obvious possibility which, while not available to Korsgaard, is available to the realist: among a set of rival normative principles that equally solve a practical problem of agency, one principle is best if it has the virtue of being *true*. The reason such a move is not available to Korsgaard is that her constructivism denies that there is anything to be said for a principle in the way of truth apart from how well it serves a practical function such as solving the basic practical problem of agency. Truth is a derivative notion, rooted in practical functions. But for the realist, a principle's truth is a separate matter from its practical function in enabling agency (as shown by the fact that even false normative principles can enable people to act on perceived normative reasons and thus exercise agency). Thus, truth can come in to help set some principles apart as superior solutions to practical problems.

Korsgaard's idea of anchoring a practical principle to the will via its role in practical problem solving turns out to be entirely compatible with realism. The practical problem we are concerned with is the agent's need for some normative principle—some conception of which ends are worth pursuing and which considerations are good reasons for act-ing—without which she would be unable to exercise rational agency. But it makes no difference as far as solving that practical problem is concerned whether the normative principle an agent adopts is the up-shot of a constructivist derivation based simply on an adequate concep-tion of the problem, with no other justification than that it solves the problem, or whether in addition it enjoys the status of being indepen-dently true. As we saw in the critique of the constructivist arguments, an agent who adopts a normative principle thereby solves her practical problem, whatever her beliefs about where the principles come from and whatever the facts may be about such things.

Suppose, then, that the formula of humanity turned out to be an evaluative and normative truth on a realist model rather than a product of Rawlsian-Kantian construction. In other words, suppose it is simply a fact to be recognized about rational beings that

P: By virtue of their rational agency, all rational beings are ends in themselves, possessing inherent and unconditional value that ought to be respected by refraining from treating them as mere means and by treating them always as ends in themselves.

The truth of P, in a realist sense, would in no way prevent P from functioning for an agent as the solution to her practical problem of needing some normative conception through which to exercise her agency. It wouldn't be merely a solution to that practical problem, since it would also be true or correct in a way that cannot be construed simply

in terms of such practical functions. But nothing prevents it from also being a successful solution to the practical problem.

If someone now asks "Why should an agent take P seriously and apply it in deliberation?" or "Why should P engage an agent's will, having practical significance for her?" the following answer is available to the realist: although P is not strictly necessary to solve the practical problem she faces as an agent—since other incompatible principles might also serve that practical function—P can reasonably be claimed to stand out from such false alternative principles as the best and indeed only acceptable solution. It is the only acceptable solution precisely because of its truth and their falsity. From the realist's point of view, P is thus not merely a normative truth spinning off by itself, lacking practical connection to the will, but is instead precisely the appropriate solution to the will's practical problem of agency.

Now this move may seem to beg the question against Korsgaard in its appeal to the relevance of P's truth to set P apart from other incompatible principles: for Korsgaard's project is rooted precisely in doubts about the role of such truth in an account of normativity. But it is one thing to argue—within a constructivist framework that has dispensed with any truth or falsity for normative principles over and above their practical functions—that normative force has nothing to do with such truth. It is quite another thing to claim that such truth or falsity would be irrelevant to normative force even if it did exist. Yet the existence of such truth and falsity is precisely what we are supposed to be imagining in considering the resources available within a realist framework. To deny the normative relevance of such truth or falsity, were it indeed part of the picture, would be distinctly implausible—even if we grant that the truth of a principle may not be sufficient for its normative force without some kind of practical anchoring.

It is important to remember that our topic all along has been normative force, not merely pragmatic value to the agent, such as enabling her to exercise agency. In judging the acceptability of a candidate solution to the practical problem of agency, we therefore have to think not only about what the principle may do for the agent but about whether it can also plausibly be regarded in the end as having genuine normative force. And while those sympathetic to Korsgaard's constructivism might reasonably try to make sense of a principle's having normative force while lacking realist truth (because it is neither true nor false in a realist sense), it would be a very different matter to suggest that a principle that is straightforwardly false nonetheless has genuine normative force, bearing on what a person unqualifiedly ought to do. That would be simply to abandon any recognizable and distinct notion of the normative over and above that of the merely pragmatic. Someone might, of course, mistake a false principle as having real normative force

if she believes it to be true, just as an agent who mistakenly believes that the glass on the table contains water (where it is actually petrol) may wrongly take this proposition to give her genuine normative reason to drink it. Indeed, in such cases of error we might even speak of the subjective normative force of a principle or proposition: for there is a sense in which relative to the agent's beliefs, she ought to take a drink from the glass or to act as the principle indicates; such a course of action might be fully rational, and it might even be irrational for her not to. But such purely subjective normativity is not what we are concerned with in this debate. We are interested in what there is genuine, unqualified normative reason for an agent to do in the familiar sense in which we understand that there is not in fact good reason for the thirsty agent to take a swig of the liquid in front of her even though she thinks there is—as reflected in the fact that a benevolent and informed advisor would tell her not to drink.

To return to the point, then, if we are speaking of genuine normative force, packing the punch of an authoritative, unqualified and not merely belief-relative "ought," then a false principle Q will lack normative force. This is true even if it has a certain pragmatic value to the agent, such as enabling her to exercise her agency by providing her with a normative scheme so that she can act on considerations she takes to be reasons. By contrast, a true principle P, while equally fulfilling this crucial practical function and solving this practical problem of agency, does so in a way that preserves the plausibility of attributing genuine normative force to it.

The reason why the truth of a principle clearly matters within a realist framework and its falsity so clearly undermines normative force is that the realist wants to link normativity to goodness of action, as described at the beginning. We are imagining supplementing this approach to normativity with the practical-problem-solving device for securing practical relevance to the agent, but the idea of goodness of action is still at the heart of the realist's picture of normativity. And only true practical principles shed light on the truth about goodness of action. A false principle can no more show that an action would be good and so ought to be done than a false theoretical principle or premise can show that an implied conclusion is true and so ought to be believed. So the realist has good reason for insisting that only true practical principles can have genuine normative force and false ones cannot. The realist therefore has exactly what is needed to single out true practical principles as the best and indeed only appropriate solutions to the practical problem of agency. This makes possible an account of normativity that combines realism with the practical-problem-solving device for anchoring normative truths practically to the will, which may be summed up as follows:

Normative force is to be understood in terms of the goodness of action—as performed by the agent in question, in the given circumstances—that is the subject matter of the true practical principles that are the only appropriate solution to the practical problem of agency that the agent has no choice but to deal with and solve.

IX. THE REALIST'S ANSWER TO KORSGAARD'S CHALLENGE

We thus have an answer to the argument against realism from Section II, concerning the "open question" about applying normative truths we come to know. The reason there might have seemed to be a significant open question about the application of an independently true normative principle was that the principle was presented simply as a theoretical truth about the normative realm, the role of which was just to be grasped and applied (or not) in deliberation. By incorporating Korsgaard's idea of practical problem solving, however, we have now arrived at something more practically grounded. A normative truth would have clear practical significance for the deliberating agent because it is at the same time nothing less than the appropriate solution to her basic practical problem: she needs to be able to view some ends and reasons as good in order to function as an agent, and a normative truth is part of the normative outlook that enables her to do this properly and without error. Far from being a removed truth of dubious relevance to her will, it is a truth precisely about what she needs to do in order appropriately to exercise her will and constitute herself as an agent.

The answer to Korsgaard's worry about an endless regress of justifying norms (Sec. II) is therefore substantially the same as her own: we needn't try to explain why one should care about and apply a principle such as the formula of humanity by appealing to further norms of application, which in turn require more such norms, and so on ad infinitum. Instead, we justify caring about and applying the principle by arguing that it appropriately solves our practical problem of agency, which puts a stop to any such regress and secures the practical relevance of the principle to the will. If P is a normative truth about fundamental human relations, this is not merely some abstract truth on a par with truths about the composition of distant planets but a truth with a practical, agency-enabling function, as part of the best solution to the fundamental problem of agency. The connection to agency, bringing out the relevance of the principle in question to the agent as an agent, is no less clear and secure just because the principle is a realist truth instead of being constructed. A principle needn't be *constructed from* the will's procedures in order to be shown to have *practical relevance for* it, and it was only a conflation of those ideas that made realism seem so problematic to begin with.

In Section I, I said that it is a condition of adequacy for an answer to the normative question about moral principles that it be something an agent in a tight corner could reasonably be expected to find compelling, at least if the relevant justifications were fully laid out. This is not to say that moral principles have binding normative force only for reasonable people, but is simply to recognize with Korsgaard that a legitimate claim of binding normative force cannot be something a reasonable person could just shrug off as irrelevant to her from the perspective of deliberation: there has to be something connecting the principle with the agent's will in the right way, showing why it ought to be of concern to her as an agent deliberating about what to do. We have now seen how the realist is able to meet this condition of adequacy. A person in a tight corner may not like the sacrifice a moral requirement demands of her, but she may reasonably be expected to recognize the normative force of that requirement for her if we have both justified the relevant principle, showing why it is true, and shown how she as an agent has no choice but to adopt some practical principle in order to function as an agent and so must in effect choose between this true principle and other incompatible and thus false ones. If reasonability requires anything, it will require choosing truth over falsehood in solving a practical problem we cannot avoid. The realist therefore has the resources to answer the normative question in a way that reasonable agents will find compelling.

Obviously, none of this guarantees that a given agent will ultimately abide by an onerous moral requirement or even that he will care enough about it to agree that he should. One might worry, then, that we have made no progress after all: a sufficiently tough-minded character may still always come back with "Why should I care?" even after we have made a compelling case for the reasonability of caring and of acting accordingly; he may, after all, care about some things more than he cares about being reasonable. Does this mean that the normative question has not ultimately been answered after all—that there is still an important "open question" calling for some kind of philosophical closure?

It does not. There was in fact never any real possibility of eliminating such a response, as through an appeal to practical necessity to gain some kind of leverage against any and every agent as such. This, however, was also never necessary. The normative question is a demand to be shown the genuine reason-giving force of moral requirements, and we are not revealed to have failed in this simply by the possibility of an agent who refuses to care about whatever we come up with, however reasonably compelling it may be as a justification. Our real challenge from the beginning was simply to find the right kind of justification, which meant finding a justification that an agent could reasonably be

expected to find compelling. The point of highlighting the hard-pressed agent facing an onerous moral requirement was not to suggest that a satisfactory answer had somehow to eliminate the possibility of indifference or to punish it with practical consequences for agency but rather just to stress the importance of showing the intrinsic relevance of moral requirements to the will. This is what Korsgaard argued has not been shown by a typical realist response to the normative question: merely claiming that a normative principle is true does not by itself show why such a truth should matter to the will. That was the challenge put to the realist.

As we have seen, however, the same challenge may be raised in connection with constructivist appeals to facts about practical consistency. In both cases, the realist and the constructivist might respond with a quick appeal to reasonability, arguing that goodness or consistency, respectively, are issues any reasonable agent would care about, as they are intrinsically normative. Korsgaard, however, wanted something deeper than this, something which would tie the principles in question to any agent's will without relying on claims of intrinsic normativity—something she thought constructivism could deliver but realism could not. This she sought to accomplish through the practical-problem-solving model of normativity, but the upshot falls short of what she had desired. First, her necessity-oriented approach to that model has apparently failed in connection with the principle at the heart of Kantian ethics. At most, she could try to make the case for this principle's being the "best" solution to the practical problem of agency, within a constructivist framework (though she would first need to overcome the other problems with the constructivist derivation). Second, even if this were successful, it would fail to achieve her goal of showing constructivism to be on firmer ground than realism. For the very same model may be exploited by the realist to explain the relevance of normative truths to the will within a realist framework, likewise securing the connection to the will by showing the principles in question to be the best solution to the practical problem of agency.

It remains to be seen whether purely constructivist derivations of ethical principles—starting with nothing but generic conditions on the exercise of agency and deriving ethical normativity from what is involved in meeting those conditions—can somehow be made to work, yielding at least interesting conclusions about commitment.[78] Likewise, it remains to be seen whether ethical realism can successfully answer the variety

---

78. The doubts I have raised about even this weaker version of such a project echo Gibbard's skepticism about what he refers to as Korsgaard's "moral logicism," i.e., her project of deriving morality from the "sheer logic of agency" (Gibbard, "Morality as Consistency in Living," 149, 152–53).

of objections that have been raised against realist views of ethical principles. What I have tried to show is that nothing in Korsgaard's work succeeds in showing that realism is generally worse off than constructivism with respect to the issue of normativity. Indeed, realists have the resources to handle the central issues at least as well as Korsgaard does, making her call for a deep practical turn in ethics premature at best. Ethics is no doubt practical insofar as "it is reflection about what to do," but this does not imply that it is therefore "not theoretical" or that it is "not reflection about what is to be found in the normative part of the world."[79] Ethics can be both practical and theoretical—a search for knowledge of normative truth for the sake of the practical end of living well—and the theoretical aspect needn't pose any obstacle to meeting the demand for practical relevance.

79. Korsgaard, *The Sources of Normativity*, 116.