THE JOURNAL OF PHILOSOPHY
VOLUME C, NO. 7, JULY 2003

COMPLIANCE, COMPLICITY, AND THE NATURE OF NONIDEAL CONDITIONS*

The question I address in this article is whether wrongdoing on the part of others can have the effect of mitigating the stringency with which moral principles apply to us. Can the fact that others behave badly make it the case that we are permitted to act in ways that would be impermissible but for such wrongdoing? I believe most of us are of two minds about this question. On the one hand, we think that it is never permissible to violate our moral standards in response to others' misconduct, because to do so is simply to "stoop to their level." Two wrongs, it would seem, just cannot make a right. On the other hand, we recognize that insisting on strict adherence to high-minded standards of conduct in the face of others' wrongdoing can be naïve at best, and at worst tragically misguided. To this extent we believe that it can be necessary, and indeed morally necessary, to fight fire with fire.

Kant is generally thought of as a purist, one who adheres unflinchingly to the first of these intuitions. There is evidence of this, of course, in his notorious response to Benjamin Constant, where Kant

*I have presented this article at the University of California/Berkeley, Stanford University, the University of San Diego Values Institute (Conference on Kantian Ethics), the University of California/Davis, and the University of California/Santa Cruz. For comments on recent versions, I am indebted to my colleagues at Stanford, especially Allen Wood, Debra Satz, Michael Bratman, and Manuel Vargas. I am also grateful to Anthony Laden, Seana Shiffrin, Jay Wallace, Christopher Kutz, Aaron James, Elijah Millgram, Sarah Buss, Connie Rosati, Tom Hill, Barbara Herman, Talbot Brewer, John Doris, and James Mahon, whose responses I have found very instructive. For comments on earlier drafts, I thank Arthur Kuflik, Tamar Gendler, Michael Stocker, and several anonymous reviewers. I am also indebted to Dmitri Petrov, who provided valuable criticism at all points along the way.

1 "On a Supposed Right to Lie from Philanthropy" (1797), included in the Practical Philosophy volume of The Cambridge Edition of the Works of Immanuel Kant, Paul Guyer and Allen Wood, eds. (New York: Cambridge, 1996), pp. 611-15/8:425-30. Unless otherwise noted, all citations of Kant will refer to this volume, with page references to this volume listed first, followed by references to the standard Academy Edition.

maintains that it is impermissible to lie to a murderer even for the sake of keeping him from his intended victim. Moreover, Kant’s biting of the bullet in this particular case seems to be motivated by a fundamental feature of his theory, namely his conception of the moral law as a categorically binding imperative. Consider the following passage from the *Groundwork*:

...every rational being must act as if he were by his maxims at all times a lawgiving member of the universal kingdom of ends.... Now, such a kingdom of ends would actually come into existence through maxims whose rule the categorical imperative prescribes to all rational beings if they were universally followed. It is true that, even though a rational being scrupulously follows this maxim himself, he cannot for that reason count upon every other to be faithful to the same maxim.... [N]evertheless that law...remains in full force because it commands categorically. And just in this lies the paradox that the mere dignity of humanity as rational nature, without any other end or advantage to be attained by it—hence respect for a mere idea—is yet to serve as an inflexible precept of the will....

To think that others’ wrongdoing can in any way mitigate the stringency with which the moral law applies to us is, presumably, to deny its categorical status. For others’ wrongdoing can only affect us by altering the causal nexus within which we act. That nexus determines the consequences of our own lawful action. But to think that the value of our own lawful action depends upon its consequences is simply to think of the law as a hypothetical imperative. Given that the moral law is categorically rather than hypothetically binding, others’ misconduct cannot undermine its normative force.

There are, however, conflicting strains in Kant’s corpus. The following passage from his *Lectures on Ethics*, for example, contains an expression of a more pragmatic position:

If we were to be at all times punctiliously truthful we might often become victims of the wickedness of others who were ready to abuse our truthfulness. If all men were well-intentioned it would not only be a duty not to lie, but no one would do so because there would be no point in it. But as men are malicious, it cannot be denied that to be punctiliously truthful is often dangerous...if I cannot save myself by maintaining silence, then my lie is a weapon of defense.

---

The image of a lie as a weapon of defense is certainly intelligible, but it is in tension with the central Kantian claim that the value of right action—and presumably the disvalue of wrong action—lies in the form of its maxim rather than its fitness to produce a certain result. How, then, can the more pragmatic position expressed in this passage be reconciled with Kant’s more familiar purism?

I. THE PROBLEM

Rather than pursue this as an exegetical question particular to Kant, I want to ask more generally what structure a nonconsequentialist moral theory has to have if it is to make sense of both purist and pragmatic intuitions with regard to the wrongdoing of others. Christine Korsgaard has attempted to do the same, most notably in her article “The Right to Lie: Kant on Dealing with Evil” (ibid.). There Korsgaard argues that Kant could have offered a limited justification of the more pragmatic position by appealing to a distinction between “ideal” and “nonideal” conditions. On this account, nonideal conditions are defined as conditions under which the attempt to abide by the moral law in its most stringent form “would make you a tool of evil,” and this consideration is supposed to justify your acting on a more permissive version of that law under such circumstances (ibid., p. 153). But exactly how is the justification supposed to work? If the sense in which you are “made a tool of evil” is merely that you are made to bring about a very bad result as a consequence of your truth-telling, then it would seem the moral value of that action depends not on its lawful form but on the amount of good or evil it brings about. Although Korsgaard would clearly reject this interpretation of her position, her own account of it is sufficiently unclear to leave the worry open. And without an alternative interpretation of what it is

4 On the one hand, Korsgaard defines nonideal conditions as those under which the attempt to conform to the moral law in its most stringent form would make you a tool of evil (p. 153). On the other hand, she seems to agree that Kant’s view seems to make the very existence of this sort of condition an impossibility. This is because, as Korsgaard explains, Kant holds a “clean hands” view of moral responsibility, according to which, as long as you do what is required of you by the moral law, you cannot be held accountable for any bad consequences which might result from your lawful action (p. 143). But this suggests that your compliance with the moral law can never in principle count as complicity with evil. If nonideal conditions are defined as those in which your compliance would count as complicity, then the clean hands view of responsibility rules out the possibility of Kantian nonideal theory. If the definition of nonideal conditions is supposed to provide a reason for mitigating the clean hands view, however, (and Korsgaard, on p. 150, suggests this is its role) then we need a different account of what the nonideality of nonideal conditions consists in. On p. 148, Korsgaard offers a gloss on Rawls’s definition of nonideal conditions as conditions which “exist when, or to the extent that, the special conception of justice cannot be realized effectively.” Characterizing the defect in this way, however, makes it sound as if the essential problem is one of inefficacy; nonideal conditions
to be made a tool of evil, the appeal to nonideal conditions can appear to be an ad hoc way of limiting the scope of Kant’s purism.

That said, I do believe the general strategy of limiting purism by appeal to some conception of nonideal conditions is promising. What remains to be seen is whether it is possible to develop a conception of what the nonideality of nonideal conditions could consist in, such that those conditions could play the mitigating role claimed for them within a thoroughly nonconsequentialist moral theory. My primary

are those under which adherence to ideal principles is somehow instrumentally defective. But the existence of this sort of defect could not provide grounds for departing from ideal principles on a truly nonconsequentialist account of such principles. Hence my dissatisfaction with the account of the nonideality of nonideal conditions as it appears in the “Right to Lie” article. That said, I believe Korsgaard implicitly revises this account, though without explicit mention of the ideal/nonideal distinction, in her later paper, “Taking the Law into Our Own Hands: Kant on the Right to Revolution,” in Reclaiming the History of Ethics: Essays for John Rawls, Andrews Reath, Barbara Herman, and Korsgaard, eds. (New York: Cambridge, 1997), pp. 297-328, to which I am greatly indebted. I believe most of what I have to say in this paper is consistent with Korsgaard’s position in this later paper, though that paper is still, I think, too vague about what the criterion is for determining whether or not a conscientious agent has reason to “take the law into his own hands” (cf. note 28).

Readers familiar with Liam Murphy’s work will recognize that I am setting up the problem in a different way than he does. Murphy starts out by defining nonideal conditions as conditions under which there is some significant amount of noncompliance, and then asking whether this condition has any bearing on the extent of our responsibilities. The answer to that question may turn out to be yes or no, and indeed Murphy’s answer is no—the responsibilities we have under conditions of partial compliance are not significantly different from the responsibilities we have under conditions of full compliance—see his Moral Demands in Nonideal Theory (New York: Oxford, 2000). By contrast, I am assuming that the concept of a nonideal condition is the concept of a mitigating condition, a condition the existence of which has the effect of undermining the stringency with which some set of moral or political principles applies to us. I am asking what has to be true of such conditions for them to play this role, and I am using the condition of partial compliance as a test case to try to answer this question because this is one of the conditions which John Rawls identifies as “nonideal” in his original use of the concept—see Rawls, A Theory of Justice (Cambridge: Harvard, 1971), especially sections 2 and 39. There are two main reasons why I think the notion of nonideality should not simply be identified with the notion of partial compliance. One is that partial compliance is only one type of nonideal condition, according to Rawls, the other being the existence of “natural limitations and historical contingencies”—A Theory of Justice, p. 246. So, the nonideality of nonideal conditions cannot consist in the existence of noncompliance. Second, Rawls’s criterion for drawing the line between the ideal condition of full compliance and the nonideal condition of partial compliance is not simply that under full compliance, no one violates the principles of justice, whereas under partial compliance, some people do. Indeed Rawls states that under full compliance, the violations which do occur “are counted as exceptions”—A Theory of Justice, p. 245. Thus, there must be some principle according to which we judge whether or not the degree or character of noncompliance is significant enough to not be discountable in this way. I take it that in order to find this principle, we need a prior account of what the nonideality of nonideal conditions consists in. I am grateful to Seana Shiffrin for urging me to be clearer on this point.
aim here is to lay out the beginnings of such an account, by examining a model of a type of rule-governed activity that would allow for an ideal/nonideal distinction of the right kind. That activity is a social practice, an activity governed by constitutive rules that apply to individuals in virtue of their occupancy of certain roles or positions. Within a practice, I will argue, the wrongdoing of others can pose a threat to us not simply by making it likely that we will, through our lawful activity, produce a bad outcome, but rather by undermining the integrity of that lawful activity. This is so, I will claim, because within a practice, our actions depend constitutively on others’ compliance with the practice, so that beyond a certain threshold, others’ wrongdoing can alter the character of what we ourselves are doing, depriving it of its status as an instance of participation in the form of activity defined by the practice rules. If this is right, then it is in this sense that nonideal conditions, at least with respect to a set of practice rules, can be defined as those under which the wrongdoing of others threatens to make us tools of evil. The further question, which I will not be able to address here, is whether action on the moral law can be construed analogously.

One additional note: apart from its implications for Kant’s moral theory, I believe the account of nonideal conditions I offer here in connection with practice rules sheds light on the possibility and structure of a certain kind of moral dilemma that Thomas Nagel has called a “blind alley.”6 A blind alley is a situation in which an ideally virtuous agent is forced, through no fault of his own, to act in a way that is insufficiently justified. Nagel believes that such a predicament is possible, because he believes we are at bottom torn by two inescapable yet irreconcilable moral standards—one deontological and the other utilitarian. On Nagel’s view, many instances of the tension

---

6 “War and Massacre,” in his collection, Mortal Questions (New York: Cambridge, 1979), pp. 53–74, here p. 74. This essay, along with Michael Walzer’s “Political Action: The Problem of Dirty Hands,” Philosophy and Public Affairs, 11 (1973): 160–80, gave rise to a significant body of literature on the question whether “moral dilemmas” are possible. For a basic anthology on the topic, see Christopher Gowans, ed., Moral Dilemmas (New York: Oxford, 1987). My claim here is to give an account of the origin of a certain kind of moral dilemma, though there may be others. One type of moral conflict I do not purport to shed light on is the kind that results from an overload of obligations. A person might experience such an overload because she occupies two distinct roles (say, that of parent and that of philosopher), or because she occupies one role to which a multitude of obligations attach (for example, University Provost). I leave it open how best to understand these types of conflicts, and whether or not they have the features of a “blind alley.” As will become clear in what follows, the type of conflict I describe results not from overload, but from corruption of the background conditions necessary for obligations to sustain their binding force.
between purism and pragmatism can be accounted for in terms of this fundamental conflict. Now part of the appeal of Nagel’s account, I take it, is that it seems to fully appreciate the depth and tenacity of certain types of moral conflict. This is in contrast to what Bernard Williams has called the “glib moralism” of theories which contend that the sense of moral conflict is always attributable to some error or shortcoming on the part of the agent. 7 I will claim, however, that an agent committed exclusively to a nonconsequentialist ideal of conduct can, under nonideal conditions, be faced with a predicament that has all the relevant characteristics of a blind alley. If this is so, then it is possible to avoid glib moralism without giving up on the unity of moral theory.

II. ACTION AS INTERACTION: THE CASE OF SOCIAL PRACTICES

II.1. Rawls on practice rules. The conception of a social practice which I will use as my starting point comes from John Rawls’s early article, “Two Concepts of Rules.” 8 There Rawls draws on a distinction, originally put forth by H.L.A. Hart, between nonconstitutive and constitutive rules. 9 In Rawls’s terminology, the distinction is between “summary” rules, which we normally call “rules of thumb,” and “practice rules,” which define games, institutions, and the like. Summary rules are generalized reports of the results of applying some other rule directly to the cases at hand. For example, the general rules employed at the middle level of a utilitarian theory are regarded as summary rules when they are thought of as approximating the results of applying the principle of utility case by case. But the use of summary rules is not particular to utilitarianism. Indeed, Kant’s duties to self and others can also be regarded as summary rules, generalizations of the results of applying the moral law to particular maxims, case by case. Practice rules, by contrast, are not arrived at by generalization, and do not function as reports of the results of some prior procedure. Rather, they define procedures compliance with which constitutes participation in some new form of activity. By complying with the rules of baseball, for example, a participant in that practice makes his behavior count as a move in that game (for example, hitting a

“home run”). Rawls’s claim is that such a move is, in the first instance, a piece of behavior that is only properly described in terms of the practice rules that govern it. Outside the “stage setting” of the game of baseball, for example, it is indeed physically possible to hit a hard, leather-covered sphere with a strangely-shaped piece of wood so that it travels a certain distance along a certain trajectory, but it is not logically possible for that bit of behavior to count as “hitting a home run.” The same is not the case with actions falling under summary rules, the proper descriptions of which have to be assumed to be fixed prior to the application of the summary rule. In this sense, Rawls maintains, practice rules are logically prior to the actions falling under them, whereas summary rules are not (op. cit., p. 37).

Rawls’s main point in “Two Concepts of Rules” is that the justification of an action falling under a practice is different from the justification of the practice itself (op. cit., p. 31). Because the actions falling under practice rules are logically constituted by those rules, such actions can only be justified by being shown to be in accordance with those rules. To justify a move by showing its conformity to some standard that is independent of the game is to justify it as some other, practice-independent form of behavior, rather than as the move that it is. Rawls applies this insight to argue that participants in practices that are governed by standards other than the principle of utility cannot, without logical incoherence, justify their actions by appeal to the principle of utility. He claims, for example, that one who occupies the role of “promisor” cannot justify keeping a promise on the grounds that doing so would promote the most good overall, because to do so would simply be to reveal that the person does not understand what promising is. The only justification of promise-keeping that he can offer, Rawls maintains, is one that appeals to the constitutive rules of promising (op. cit., p. 39). Moreover if he wants to justify breaking a promise, he can only do so by appealing to one of the various excusing conditions built into the content of the practice rules themselves (op. cit., p. 32).

Although the justification of particular actions falling under practice rules cannot make appeal to practice-independent standards, Rawls maintains that the justification of a practice as a whole must appeal to such standards. For if there is any reason to set up a practice in the first place, Rawls claims, that reason must be based on a principle that is defined prior to the practice. Hence the standpoint of what I will call the social engineer is one that can be governed directly by the principle of utility, even though the standpoint of the participant
cannot (op. cit., p. 31). By the same token, a participant who wishes to make changes in the practice as a whole cannot do so without stepping out of his role as a participant and instead acting from the standpoint of a social engineer. Hence one cannot justify an act of, say, promise-breaking on the grounds that the rules of promising are not serving their practice-independent purpose. That can be a justification for changing the rules, but it cannot be a justification for breaking them.

II.2. When a practice becomes a sham. Rawls’s aim in making this logical point is to strengthen the case for a two-level version of utilitarianism. His claim is that if general rules are construed as practice rather than summary rules, utilitarianism need not permit counterintuitive violations of our ordinary standards, violations such as punishing the innocent for the sake of the greater good. Later on, I will have something to say about whether this defense of utilitarianism is successful. But for now I am simply going to draw on what I take to be Rawls’s valid and illuminating analysis of practice rules in order to show—or rather to begin to show—how it can provide support for two-level Kantianism. As such, my strategy is to put Rawls’s notion of practice rules to a use that is, in a sense, directly opposed to the use he originally made of it in “Two Concepts of Rules.” Instead of arguing that the notion of a practice rule can be invoked to build the right kind of fixity into a two-level utilitarian theory, I will argue (or rather, provide the first step of an argument) that that notion can be invoked to build the right kind of flexibility into a two-level Kantian theory.

Rawls’s argument focuses on examples where an agent fails to follow the rules of a given practice, choosing her action instead on the basis of some practice-independent standard, while still attempting to justify what she does by appeal to the practice rules. His point is that this results in a kind of constitutive and justificatory failure—having chosen the action on the basis of practice-independent principles, the action she in fact performs is not properly described as a move in the practice, and hence the practice rules cannot be invoked to justify it. The cases I will be focusing on have the same structure, except that the source of constitutive and justificatory failure is external to the agent’s will. Suppose you are a conscientious participant in a practice, and you choose to perform a certain action on the basis of

---

10 Rawls does not explicitly use the concept of a “social engineer.” I am using this term so as to be able, later, to draw a distinction between the standpoint of the “social engineer” and that of the “reformer.”

11 This claim will require qualification later, when I discuss the role of the “reformer.”
your judgment that doing so is in conformity with the practice rules. Suppose further that your judgment is correct—the practice rules do demand this sort of conduct in this sort of situation. Moreover, nothing prevents you from acting on your decision, so you go ahead and do so. Have you thereby made the move?

While it may seem obvious that you have, this is not necessarily the case. Consider a simple example. You are a party to a dispute, and you are negotiating with the other party. Let us assume that negotiation is a practice, and that it is constitutive of this practice that actions falling under it be regulated by an ideal of “good faith”; that is, one who is not at least approximating this regulative ideal is not, strictly speaking, negotiating. Let us assume further that you are indeed negotiating in good faith: you put forth your claims as reasonably as you can, and you try to listen to the considerations of the other side as impartially as you can, in the interest of reaching a mutually acceptable agreement. During the course of the negotiation, however, it becomes clear to you that the representative of the other side is not governing himself the way you are. What he wants to do is to give the appearance that he is negotiating, while stalling for time. Outwardly, he goes through the motions of listening and responding to your arguments, but you are aware that he is really just trying to manage you, to work around you like an obstacle, without really engaging with the substance of your claims. Given this situation, what is the proper way to describe your own actions? Are you negotiating, or are you just babbling on?

If it makes sense to raise this question, this suggests that we take there to be some threshold beyond which the noncompliance of another can have a bearing on the integrity, and not merely the efficiency, of your rule-governed action. Let us say that an action is productively successful when it efficiently promotes the end it is designed to achieve, whereas an action is constitutively successful when it fulfills the necessary and sufficient conditions for being an instance of the type of activity it is supposed to be. Now, it is clear that the noncompliance of others can have the effect of undermining the productive success of your actions. This is true, moreover, whether or not the

---

12 By “action,” I mean to refer not only to external acts but to deliberative and volitional activity as well. In general it is important, when using the practical model, to guard against an overly legalistic construal of practice rules. It is clear, for example, that Rawls takes the rules of judicial punishment to constrain not only a judge’s external actions, but also his internal actions of deliberation and choice. I believe this is because Rawls takes the actions falling under practice rules to constitute forms of active participation in the practice. I argue for this in my paper, “Three Conceptions of Action in Moral Theory,” Noûs, xxxv, 1 (2001): 93–117, and I make that argument in a different way below, section 11.3.
end to be achieved is practice-dependent or practice-independent. That is, the conduct of others can undermine your productive success by interfering with your ability to promote ends that you must adopt in virtue of being a participant in the practice, as well as ends you adopt independently of that role. In this case, your interlocutor's noncompliance makes it quite unlikely that you will achieve your practice-dependent end of coming to a negotiated agreement, and he will also probably make it unlikely that you will achieve your practice-independent end of protecting the interests that are at stake in the dispute. But such productive failure is certainly compatible with constitutive success. Indeed, in some cases it will be appropriate to describe the two of you as engaging in negotiation while still failing to reach an agreement or to protect the relevant interests. The question at hand is whether it is possible for the noncompliance of one party, at least beyond a certain threshold, to make it the case that the other is not even participating in the activity defined by the practice rules, despite the fact that the latter party is following the letter of the law.

In one sense, it is trivial that such constitutive interdependence exists. Every action, whether governed by a practice or not, depends to some extent upon the presence of a constitutive background if it is to be described in one way or another, and to the extent that such background conditions make reference to the behavior of others, our actions depend constitutively on those of others. For example, in order for my action to be describable as "giving you a hard time," you have to be annoyed or inconvenienced by what I am doing. But the absence of such conditions, it might be objected, does not always amount to a constitutive failure of any kind. For this to be the case, there has to be something my action ought to be, some character it is supposed to have but it lacks. In the negotiation example, it is only the case that you are merely babbling on with reference to some standard according to which you ought to have been negotiating. But, the objection goes, in this example you are simply mistaken about which rules are in effect. You are attempting to play the "negotiation game," while your interlocutor is attempting to play, say, the "domina-

---

13 Whether the converse is true depends upon whether the end in question is practice-dependent or practice-independent. It is certainly not possible to win a baseball game if you are not playing, and in general if the end is a practice-dependent one like "winning," then constitutive success is a condition of productive success. But if the end is conceived as one for the sake of which an agent chose to enter into the practice—for example, to get some exercise—then productive success may not depend upon constitutive success. One might try to play baseball with a crew of five-year olds and end up getting a lot of exercise even if the group never succeeds in really playing the game.
tion game.” As such, there is no established practice, and neither you nor he can be described as subject to the rules of negotiation in the first place.

In some cases this will indeed be an appropriate description of the situation. But the possibility I am inquiring into is one that presupposes that the practice is sufficiently well established (perhaps in part by means of external, legalistic criteria, but perhaps by means of implicit and explicit signs of commitment communicated by each party) for you and those with whom you interact to count as occupying roles in it and as being subject to its rules. In these cases, the noncompliant party is tacitly or implicitly claiming the protections and prerogatives attached to his role while at the same time failing to live up to its demands. If, under these conditions, it is in principle possible for such noncompliance to make it the case that you are babbling on rather than negotiating, this would show that your action depends on his, not only to be describable in one way or another, but to be describable in the way it ought to be described.

I believe the possibility of this sort of constitutive failure is presupposed by our ordinary use of the notion of a “sham.” “This negotiation is a sham,” you might say, pounding your fist on the table. To make this claim is to insist that the activity in which you are engaged is no longer what it ought to be, and that this has come about through no fault of your own, but rather through the misconduct of others. Moreover you would normally make this charge as a way of calling into question the value of continuing to conform to the practice rules, given that your rule-governed actions have to some extent lost their integrity. What would make this sort of challenge intelligible? Given the constitutive nature of practice rules, the idea would have to be that conformity with the practice rules is no longer sufficient to insure constitutive success; that is, the fact that your action is chosen in accordance with the practice rules is no longer sufficient to make it count as participation in the form of activity that is, ideally, defined by those rules. Hence, when a practice has become a sham, adherence to the formal procedures in terms of which the activity is defined can take on the character of a mere formality, an empty gesture.

Aristotle wrote that a finger, when detached from the living organism of which it was a part, becomes a finger “in name only.” Similarly, I want to claim, lawfulness, when detached from the background

14 Or through some other force external to your will. Here I am focusing on noncompliance as a potential source of nonideality, but as I mention in note 5, there are others.
conditions in virtue of which its constitutive function is defined, can become merely nominal lawfulness—a kind of rectitude that is lacking, at least to some degree, in real value. I am going to call this the sound and fury problem, because it picks out a way in which forces beyond our control can deprive our well-chosen conduct of its proper significance. Now notice that if the sound and fury problem is a real possibility, it provides a model of what Nagel calls a “blind alley.” As I mentioned earlier, a “blind alley” is a situation in which an ideally virtuous agent is somehow forced to act without justification. Now, if you are faced with the sound and fury problem, you are by hypothesis the occupant of a certain role, which means you are bound to govern yourself in accordance with a certain system of constitutive rules. But given that the practice is a sham, by so governing yourself, you end up engaging in constitutively failed action. If Rawls is right in claiming that the justification of an action falling under a practice can only appeal to the practice rules, then this constitutive failure amounts to a justificatory failure as well. For when you appeal to the practice rules for justification, you are attempting to justify your action as a form of participation in the activity defined by the practice. But when what you actually do amounts to something less than this—when, instead of making the moves, you end up merely going through the motions—then the practice rules cannot justify what you actually end up doing. Can you therefore appeal to some practice-independent standard in order to justify what you actually do? Not if, as was assumed, you are correct in taking yourself to occupy the role in question, and are not simply imagining yourself to do so. Unless you decide to abandon that role (which in some cases may not even be an option), you are bound by its rules. But unless your coparticipants change their ways, those rules will be constitutively and justificatorily defective.

To summarize: because you are a participant, you have to play, and because the rules are constitutive of participation, the only way to play is to play by the rules. But because the background conditions presupposed by such rules are ill established, playing by the rules fails to amount to participation in the relevant form of activity. As such, what is a conscientious player to do? If you comply with the letter of the law, you will betray its spirit, in the sense that you will not be engaging in the form of activity in terms of which you value yourself and your conduct as a player. If you violate the letter of that law, however, you will likewise fail to participate in that form of activity, because there are no other rules in terms of which that activity is defined.

II.3. How a practice becomes a sham. At this point an objection arises. The problem as I have just described it presupposes that there can
be a gap between adherence to the letter of the practice rules and full-blooded participation in the relevant form of activity. But if the rules are truly constitutive of the activity, how can any such gap exist? The idea that there can be such a gap seems to presuppose that we have an independent idea of what the activity ought to be like. But where does this idea come from? Without an answer to this question, the account I have offered is vulnerable to Nagel’s suspicion that the standard to which we (perhaps unconsciously) appeal when we call a given practice a sham is none other than the principle of utility.\footnote{The same suspicion could of course be shared by a thoroughgoing utilitarian like Henry Sidgwick.} That is, we call a practice a sham when it has been rendered inefficient, or sufficiently inefficient, in achieving the practice-independent ends it was originally designed to promote. Above that threshold, we can afford to act like purists, abiding by the rules regardless of the consequences. But below that threshold, such purism becomes too costly, and we are justified in acting more pragmatically. And even if it is the case that when we act more pragmatically our actions have to be described somewhat differently—not as moves but as “shmoves”—so what? The point is to adopt the most efficient method available, given the circumstances, to advance certain practice-independent ends. Indeed, the efficiency of the practice is ultimately what justifies our taking up roles within it, so its inefficiency can justify our abandoning those rules in particular cases where adhering to them would be too costly.

Of course, this is the sort of position to which Rawls was originally responding in “Two Concepts of Rules,” namely, a utilitarian account of practices that allows participants to take exception to practice rules on the basis of direct appeal to the principle of utility. The objection reveals that Rawls’s position was not sufficiently defended, and to the extent that it was not, my own use of Rawls is likewise vulnerable. Here is what I think Rawls needs to complete the argument. Although it is the case that practices as wholes are adopted in virtue of their fitness to promote certain practice-independent ends, this efficacy is not the source of their normativity as practices, or of their bindingness upon us as participants in practices. The relevant contrast here is between a practice and a method. A method, as I think Rawls should conceive it, is a procedure that works like a tool; it is designed to serve some end, and that end is conceived as having been set in advance by a will that is constituted independently of the procedure. A practice, on the other hand, is a procedure by which a new, shared
will, along with its ends, is constituted. This might sound counterintuitive, because as Rawls pointed out, we adopt practices as wholes for the sake of achieving certain practice-independent ends, and not obviously for the sake of constituting shared wills. Even if the initial incentive to set up a practice comes from the need to achieve a practice-independent end, however, the decision to achieve that end by means of a practice rather than a method amounts to a decision to claim that end as a common end, and to come together as a community in order to realize it. Hence we might say that, whereas a method defines a way of solving a problem, only a practice defines a way of making a problem, along with its solution, count as ours.

I think this idea underlies Rawls’s account of the logical structure of practices. Rawls says that a practice is “the specification of a new form of activity.” But he is likewise committed to the view that a practice specifies a new source of agency, one to which that new form of activity is attributable. To see this, notice that actions falling under practices are attributable to agents not as particular individuals, but as occupants of roles or positions that are defined relationally in terms of one another. Such actions are thus attributable to individual agents only insofar as those agents are regarded as constituents of a shared source of agency. In this sense, the actions of each are the actions of all. They are essentially public actions, expressions of something like a shared will. If actions governed by practice rules are to be valued only as public actions, then it would make sense to say that they are only properly described as such, in terms of the rules that constitute their public character.

Of course, this notion of a shared will is one that requires elucidation, and what I have to offer here is admittedly incomplete. There

---

17 For a closely related distinction, see Anthony Laden’s instructive discussion of the rational and the reasonable, in his Reasonably Radical (Ithaca: Cornell, 2001), chapter 4.  
18 “Two Concepts of Rules,” p. 36.  
19 This may seem counterintuitive in cases where the practice involves direct competition between opposing players or teams. When a pitcher for the Red Sox strikes out a batter for the Yankees, how is this action in any way attributable to the agency of all the players involved in the game? The answer is that every time a player acts within his role as a baseball player in a particular situation (for example, occupying a particular position relative to other players), he upholds the establishment of the practice. As such he contributes to the constitution of the practice-based activities of all involved. The pitcher for the Red Sox may thwart the Yankee batter’s attempt to promote the specific practice-dependent end assigned to him in virtue of his position as a batter; but insofar as he acts well as a pitcher, he upholds the Yankee batter’s status as a baseball player engaged in the practice of baseball. Another way to see this is to notice that when the competition takes place within a fundamentally cooperative context, playing to win is conceived as an aim that is fully compatible with good sportsmanship.
is a growing body of literature on the subject of shared agency, within which there is debate about what needs to be the case in order for it to be true that “we” are acting together. Some philosophers see the “we” as a structure of interlocking individual intentions to perform the activity, and some see it as a plural subject constituted by normative claims which individuals make on one another with respect to the performance of the activity.20 For the purpose of this article, I need not take a stand on shared agency in general. However I am assuming, without argument, that at least in the specific case of a social practice, the “we” takes the latter form. In a practice, actions are attributable to a shared will because and insofar as participants make reciprocally-binding claims upon one another to adhere to the rules of the practice. This may make it look like I am employing a moralized notion of a practice, one which applies only to good or justifiable practices. The normative element being presupposed here is minimal, however. As I am understanding it, a practice must be characterized by reciprocity, but it need not be characterized by the highest degree of reciprocity. What is essential to a practice is that it is a system of social cooperation rather than a system of social manipulation; but this leaves room for practices that are less than fully democratic. For example, the institution of the Catholic Church defines a practice even though the decision-making structure of that institution is hierarchical. It is a practice because (or insofar as) it constitutes a system of social cooperation, rather than one based on coercion and force. It is cooperative in virtue of several features, which I put forward in a provisional way, and which are not intended to be exhaustive or conclusively formulated.21 First, its rules are viewed by all participants not simply as commands backed by sanctions, but as genuine obligations. Moreover, participants view themselves and each other as responsible agents capable of following those rules on the basis of their own consciences. In addition, the system of rules is oriented towards the realization of a common good, a good the instantiation of which serves to benefit each member of the community. These conditions indicate the neighborhood in which the line between practices and nonpractices is to be drawn, though there will of course be borderline cases.22


21 Here I am taking my lead from Rawls’s discussion of a “decent consultation hierarchy” in his work on international relations, entitled *The Law of Peoples* (Cambridge: Harvard, 1999), see especially pp. 71–78.

22 A further complication is that there will be cases in which an institution that has the internal structure of a practice relates to outsiders in a fundamentally uncoop-
The significance of the method/practice distinction is that it allows us to specify a nonconsequentialist basis for departing from the practice rules. For although the normativity of a practice qua method lies in its efficacy in promoting an end, its normativity qua practice lies in its fitness to constitute a shared will, a "we" to which the ends pursued, along with the moves themselves, are attributable. This is what the integrity of a practice consists in, and as long as it is intact, action in conformity with the rules will likewise have integrity as an expression of a public will, even when it fails to achieve its intended results. Under these circumstances, noncompliance is regarded as something accidental to the substance of the practice, something that therefore has no bearing on the stringency with which the practice rules bind participants. Now if all misconduct were properly regarded this way, then it would be impossible for noncompliance to render a practice a sham. That possibility rests on the assumption that at least some forms of misconduct can undermine what is essential to the practice, namely its constitutive function. Call this form of noncompliance subversion as opposed to mere transgression. The concept of subversion, I take it, is that of a form of noncompliance that has the effect of detaching the law from the constitutive background in virtue of which lawful actions succeed in being expressions of a public will. The subversive is one who not only breaks but usurps the law, making it into a tool of his unilateral will. In doing so, he makes lawfulness into a mechanism by which his private will is served, rather than an ideal standard in virtue of which a public will is constituted and expressed.

The essential threat posed by subversion, then, is that it makes you end up serving a unilateral will, and not that it makes you end up promoting an end that is bad or evil. Consider, for example, a

---

23 Korsgaard marks an analogous threshold between mere "imperfection" and "perversion" in "Taking the Law into Our Own Hands," p. 318. She does not explain its basis, however.

24 On the negative point, that what is at issue is not, fundamentally, the goodness or badness of your coparticipant's aim, my account is in agreement with the one Korsgaard gives in "The Right to Lie." However, (1) I do not claim at this point to have made an argument that applies directly to the case of practice-independent actions, like truth-telling, and (2) I believe my positive account of how you can be made a tool of evil, at least in the context of a practice, is more explicit than the account Korsgaard offers in that article. What I say here, however, is much more closely in accord with the position Korsgaard takes in her later article, "Taking the Law into Our Own Hands." Because that article deals with a problem concerning one's participation in a political institution, its relevance to the cases I am dealing
modification of the negotiation case in which your interlocutor is trying to promote the very practice-independent end in virtue of which you decided to engage in negotiation—namely that of protecting your interests. Imagine that your negotiating partner genuinely wants to find a resolution that adequately addresses your interests. But imagine that he does not want to present this resolution as the result of a process of negotiation, because he does not want to give the appearance that he has been responsive to your direct appeals. Such a person might purport to negotiate with you, while using the activity merely as a method for gathering information. Although he listens to what you say, he does not make a good faith attempt to come to a negotiated agreement with you on the basis of that interaction. His aim is to find out what you need and to develop a way of meeting your interests that he can later implement on his own, as a way, perhaps, of validating his conception of himself as a benevolent soul.

Despite the fact that the outcome at which this person is aiming is the same as the one you might have wanted to achieve through negotiation, my contention is that this sort of conduct can, at least at some point, render the negotiation a sham. For in this case your conegotiator is indeed using the activity as if it were a mere method, responding to you as a source of information but not as a source of reciprocally-binding claims. And this can put you in the same predicament of not being able to determine exactly what you are doing by acting in good faith in the face of such conduct.

This same point, that what makes noncompliance subversive is its unilateral nature rather than the value of its end, implies that the sound and fury problem can arise even in cases where the practice as a whole is oriented towards morally evil ends, and subversion is undertaken for the sake of the morally good. Consider a modification of the previous scenario. Suppose you are a middle-level Mafioso, and you have come to consult with the Godfather. The family code requires that you engage in consultation in this case—one in which, say, you need the Godfather's permission to take over a new market. Because the internal organization of the family is that of a practice, you come to the Godfather out of a sense of duty, and not simply out of fear

with here is obviously more direct. But there are other important differences between this later article and the earlier one. In the earlier one, for example, Korsgaard suggests that appeal to the Formula of Universal Law can give direct guidance to the person faced by the murderer at the door. In the revolution article, she denies that universalization is of any help to the revolutionary (p. 319). I believe my position, once it is extended to Kantian theory generally, will be in accord with Korsgaard's later view; I should emphasize again, however, that I do not claim to have yet made that extension.
that you will be punished if you do not. But now you notice a bump under his shirt, and you become aware, to your horror, that he is wearing a wire. Instead of consulting with you, he is using the consultation as a way of gathering evidence for the FBI. Perhaps his motive for doing this is self-interest; he has concluded that sooner or later he will be convicted, and has decided to cut a deal with the authorities so as to insure a lighter sentence. Or perhaps he has undergone a personal moral conversion, and he feels called upon to help the authorities bring his “family” members to justice for its own sake. In the former case, his noncompliant action is morally good, albeit only externally so. In the latter case, his action is morally good in the full sense of being the right action chosen for the right reason. From your point of view as a Mafioso, however, the moral status of his action makes no difference. In either case, the Godfather is acting on the basis of principles that are practice-independent, and to this extent he has constituted his will as a private one, while still attempting to make use of the public procedures of consultation. Insofar as you are committed to being a good Mafioso, his moral conviction operates as a subversive influence.

The type of case that might seem hardest to account for on the view I am proposing is one in which the practice is morally good and has integrity, but it fails to promote the practice-dependent ends that all participants, qua participants, are committed to realizing. To take a familiar example, let us suppose the judicial system is intact, but because of the inevitable imperfections inherent in any such system, a guilty person goes free, or an innocent person is condemned. A good participant will of course deeply regret the outcome and perhaps try to work toward improving the practice so as to make this sort of result less likely in the future. If the normativity of the practice, however, does stem from its constitutive function, and if that function is intact, this need to change the rules does not provide a justification for breaking them. That said, it might be argued that there must be some degree of inefficiency so egregious that it would justify bending the practice rules, at least where the end in question is practice-dependent.25 On the account I am proposing, however, the egregiousness of the inefficiency is only relevant insofar as it serves as an indication that the public character of the procedure is not, in fact, intact. In cases of jury nullification, for example, it is the law’s integrity as an expression of the general will that is at stake. And to establish that the integrity of that law has been undermined, one would have

25 I again thank Shiffrin for prompting me to address this sort of objection.
to argue that the law has been made to serve an alien will. When this
is the case, the egregiously bad outcome is to be avoided not because
it is an imperfect realization of a public aim, but because it is the
successful realization of a private one.

II.4. An illustration. To make my rather abstract claims more vivid,
here is a more detailed illustration of the predicament I am calling
a “blind alley,” albeit one that involves subversion for the sake of a
morally evil end. I offer it with the caveat that nothing in my argument
turns on your agreeing that this particular case counts as one in which
the practice rules have indeed been subverted. All I need is your
agreement that the example is illustrative of the sort of features that
would characterize such a condition. The example is taken from the
film *L. A. Confidential.* Among other things, this is a story about a
rookie cop who vows to play clean despite the fact that everyone else
plays dirty, only to find that playing clean amounts to violating the
very ideal of justice he set out to uphold. The climactic scene comes
in a confrontation at the end between the cop, Edmund Exley, and
the corrupt Chief of Police, Captain Dudley. Dudley has been using
his influence as Chief of Police to gain control of a drug trade formerly
run by the Mafia. In an effort to secure his power, he has framed
innocent civilians and has tortured and murdered fellow police offi-
cers. Despite an extensive cover-up operation, Exley has discovered
the truth, and Dudley stages an ambush in which he tries to have
Exley murdered. After a bloody shootout, Exley ends up alone with
Dudley, holding the Captain at gunpoint. Police sirens are heard
approaching from the distance. Dudley says to Exley, “Well, kid—are
you going to shoot me, or arrest me?” When Exley does not shoot,
Dudley says, “Always the politician, eh? Good boy. Just let me do the
talking and you’ll be Chief of Detectives.” The Captain walks out to
face the approaching police cars, leaving the armed Exley behind
him. Then he pulls out his badge and holds it up, announcing to
Exley, “Hold up your badge, so that they know you’re a police officer.”
At that point Exley shoots Dudley in the back. Then Exley goes out
to face the cars, holding his badge in the air.

The scene poses the question vividly: in this situation, which course

26 The film is based on the novel of the same name by James Ellroy (New York:
Warner, 1990); film released by Warner Brothers, screenplay by Brian Helgeland
and Curtis Hanson. The scene I describe occurs only in the film version.
27 Actually, the motives of the clean cop are not quite that simple—at the outset
of the story, he is driven more by a desire to establish a reputation as a clean cop
than by a desire to be one. But I take it that in the scene I describe, which comes
at the end, he acts on the latter motive.
of action—shooting or arresting—is demanded by the practice of policing, and by Exley’s ideal of himself as a good policeman? If there is any justification for Exley’s decision to shoot, it cannot simply be that, in this case, it was likely that the Captain would have eluded prosecution. A good cop is not entitled to choose between arresting and shooting a suspected lawbreaker on the basis of an estimation of the likelihood that he will be brought to trial and convicted. This is so even if the crimes involved are egregious, and if the likelihood of his repeating the offense if allowed to go free is high. And this is so even in light of the fact that the whole purpose of establishing a practice of law enforcement is to provide a mechanism for restraining and punishing those who break laws. Granted, if law enforcement were merely a method for achieving this goal, Exley might be straightforwardly justified in doing whatever is necessary to stop Dudley. But since law enforcement is a practice, or perhaps a subpractice within the general practice of citizenship, acts of law enforcement must, first and foremost, be public acts; they must be undertaken in such a way as to represent the united will of the citizenry as a whole, a will that is embodied more specifically in the standards and procedures of policing. As such, it looks like a good cop has no choice but to make the arrest. For if he shoots, he acts unilaterally, on his own private will, and he becomes a murderer just like Dudley.

And yet we are supposed to see Exley as being faced with a real dilemma (and most of us, by the end of the story, do feel a sense of dilemma). How can this be explained? One option is to see Exley as being torn between his conception of himself as a good cop and some other conception of himself that carries with it an obligation to promote utility in general (or to prevent suffering in general). This is the Nagelian, pluralist view, according to which our sense of moral dilemma in extreme situations like this stems from the fact that we are torn between deontological and utilitarian perspectives. But this reading of the conflict seems strained. Exley is torn precisely because he feels obligated to do what is demanded of him as a good policeman; indeed after he shoots Dudley, he holds his badge in the air, as if to show that he is (still) a good policeman. Hence I take it that in the decision between arresting and shooting, what is at stake for him is his integrity as a police officer, independent of any further obligation to promote the overall good. But if this is indeed Exley’s concern, then why is it not obvious that he ought to make the arrest? What could tempt a conscientious cop to violate the constitutive rules of policing?

The argument I have been developing suggests an alternative interpretation. It is open to Exley to ask himself whether the Captain’s criminal activity is of a nature as to undermine the public character
of the practice, and hence that of Exley’s own lawful conduct. Has the Captain merely transgressed the law, or has he subverted it? The Captain’s last words and gestures are, I take it, supposed to suggest the latter: “Hold up your badge, so that they know you’re a police officer.” The image suggests, albeit metaphorically, that the Captain is not merely misapplying the concept of a policeman to himself—he is trying to change the function of that concept, to make it stand for all he stands for—namely, the unconstrained pursuit of personal power.28 This indicates that from Exley’s point of view, the Captain is putting himself above the law in such a way as to alter its constitutive function, making it a tool of his unilateral will rather than a procedure by which a public will is constituted and expressed. If this is indeed how Exley views his situation, then he can rightly wonder whether making the arrest, though formally a public act, amounts substantively to anything more than submission to the Captain’s private will. And by the same token, he can rightly wonder whether shooting the Captain, though formally an act of murder, might in some sense amount to a surrogate for the public act of law enforcement that he must—but is not in a position to—carry out.

My argument does not require that we agree that Exley is right in thinking that the situation he faces in fact meets this description. Moreover, even if we agree with his assessment, we need not agree that the particular form of lawbreaking he ultimately opts for—namely shooting the Captain—is the least unlawful of unlawful alternatives open to him.29 What matters is that if Exley sees his situation the way

28 Korsgaard argues that what Kant finds horrifying about one particular unjust act, the formal execution of a monarch, is that “it presents an evil act in the outward form of a lawful one,” and is in this respect “like an act of a malevolent [diabolical] will”—“Taking the Law into Our Own Hands,” p. 318. Captain Dudley’s holding up of his police badge has the same property. But there is still a question whether being faced with this particular kind of horror gives us license to act in ways that would otherwise be disallowed. Korsgaard suggests an answer along the lines I am proposing here when she says, “[C]oncern for human rights leads the virtuous person to accept the authority of the law, but in such circumstances adherence to the law will lead her to support institutions that systematically violate human rights” (p. 318). I am trying to give a more detailed and more explicitly nonconsequentialist account of what this problem is, such that it can weaken the force with which the agent is bound by the letter of the law. On my account, it is only to the extent that the law has been made into the tool of another’s will that you can, in virtue of your lawfulness, be made a tool of another’s will.

29 Although the account I have been offering here is independent of Kant’s theory, it does seem worth mentioning Kant’s insistence that he “cannot deny all respect to even a vicious man as a human being; I cannot withdraw at least the respect that belongs to him in his quality as a human being, even though by his deeds he makes himself unworthy of it”—Metaphysics of Morals, p. 580/6:463. On Kant’s view, one can never become so wicked as to lose one’s nominal status as a human being. Just what this claim ends up ruling in and out is a further question. Kant took this claim
I have described, he will indeed find himself in a predicament with
the structure of a blind alley. As a good policeman, he is required to
refrain from vigilantism; he must perform only those acts that can
count as public acts, even at the cost of allowing a guilty and perhaps
dangerous person to go free. And yet he takes the situation to be one
in which the law itself is being held hostage; it is as if the Captain,
through an act of moral ventriloquism, has substituted his own voice
for the voice of the people, while leaving the outward procedures of
public expression in place. If this is so, then Exley’s adherence to those
formal procedures is at risk of amounting to an empty formality. But
departure from them is not without risk as well. For those procedures
are the only ones that have any claim at all to express the public will,
so that departure from them is not, by any publicly-communicable
standard, a way of embodying that will.

Let me be clear about what I am not claiming here. I am not
claiming that every instance of noncompliance on the part of one
or more parties renders a practice a sham. Practices are inevitably
imperfect realizations of the ideals that govern them, and it is in
general incumbent on us to abstract from these imperfections in
conducting our ordinary affairs. What I am claiming, however, is that
if noncompliance can in principle render a practice a sham, in the
normatively significant sense that I am outlining here, this is how we
have to conceive of it doing so. There is, no doubt, a threshold
judgment that needs to be made in these cases, and what I am sug-
gest ing is that there is a nonconsequentialist standard to be appealed
to in making that judgment. That standard is one which assesses the
integrity of the practice as a practice, rather than its efficacy as a
method, and hence it involves a judgment about how you would be
relating yourself to the other participants in the practice were you to
abide by the letter of a law that is being misused. It is entirely possible
that the threshold set by this standard turns out to be extremely high.

to justify a prohibition against dehumanizing forms of punishment, and yet he was
a proponent of capital punishment in certain cases. Presumably preservation of one’s
nominal status as a human being does not rule out the possibility that the claims
one makes on others in virtue of that status might be weakened, if one has acted in
a way that is essentially incompatible with that status. In such a case, we might say
that one’s status as a human being has become merely nominal.

30 This is not to deny that there can sometimes be value in adhering to a formal
procedure even when it has, to some extent, become an empty formality. One might
adhere to the law as an “aspirational” gesture, a way of inviting the other back into
the moral community, and as an expression of hope that the proper function of the
law might be restored. My view implies only that to adhere to the law in this way is
somewhat different from adhering to it under conditions where its constitutive func-
tion is unimpaired. Thanks to Shiffrin for prompting me to address this.
It may be that there are very few cases in which noncompliance has this subversive character, in which case there would very seldom be grounds for departing from strict adherence to the letter of the practice rules. But to deny that any such threshold exists in principle, I maintain, is tantamount to claiming that participation in a practice is a kind of blind proceduralism. If what is valued is adherence to the letter of the practice rules as such, independent of any deeper conception of what one is thereby doing, then practice rules might indeed be incorruptible. Noncompliance would not threaten their constitutive function, because they would not have one. But the price of such incorruptibility would be to render them pointless. The point of setting up a practice is to establish a way to act as one in relation to some problem or other. The normativity of practice rules depends upon their playing this constitutive, unifying role. Hence, we can at least conceive of those rules losing their normative grip should they be forced to serve some alien function. The noncompliance of participants is at least one potentially corrupting influence on a practice—one type of nonideal condition—though there may be others.31 What is important is that this construal of what nonideality consists in provides a nonconsequentialist basis for the claim that the wrongdoing of others can, at least within the context of a social practice, mitigate the stringency with which otherwise unexceptionable rules apply to us.

III. IMPLICATIONS

III.1. How this account sheds light on the possibility of moral dilemmas. This analysis shows, first of all, that a blind alley can arise without supposing that the agent is torn between his practice-dependent obligations on the one hand and utilitarian considerations on the other. It arises simply in virtue of his commitment to the practice, combined with the fact that the subversive activity of another has made the practice constitutively defective. But this account can also explain why action under nonideal conditions can appear to be more goal oriented than action under ideal conditions, and hence why it can look superficially like action governed by consequentialist reasoning. The explanation is as follows. Participation in a practice is not blind rule-following; it is a self-conscious form of activity in which an agent sees himself as

---

31 Here I am referring to the other main part of nonideal theory as it is described by Rawls, the part which deals with problems arising out of “natural limitations and historical contingencies,” as opposed to those arising out of the noncompliance of others—see A Theory of Justice, p. 246. I have tried to argue that human immaturity constitutes one kind of nonideal condition in my “What Is a Child?” Ethics, cix (July 1999): 715–38.
relating to others as a coparticipant in virtue of his adherence to the rules. As such, every participant must be conceived as adhering to the rules on the basis of a background assessment of the integrity of the practice as a whole. Under nonideal conditions, this background assessment comes to the fore, because the practice is judged to be in need of repair. This can seem to force the participant into the position of the social engineer, since the social engineer is one who takes the functioning of the practice as a whole as his object. But the similarity is superficial. To distinguish the standpoint of the social engineer from the standpoint assumed by a participant under nonideal conditions, I will call the latter the perspective of the reformer. The reformer differs from the social engineer in two crucial respects: (1) the reformer takes himself to be a participant in the practice, and hence to be bound by the practice (construed broadly) in every action, whereas the social engineer, on the other hand, does not think of himself as a participant; and (2) the reformer adopts the end of bringing the practice into conformity with its internal constitutive standard, whereas the social engineer attempts to improve its efficiency relative to an external standard.

As for (1), because a reformer is simply a participant under nonideal conditions, a reformer is first and foremost committed to acting as far as possible in accordance with the demands of the practice, albeit in a situation where this may or may not mean acting according to the “letter” of its law. It is only provided this procedural condition is fulfilled that the reformer is entitled to work toward bringing about any specific state of affairs. As for (2), the reformer is a participant who, in the face of nonideal conditions, may find it necessary (for the sake of preserving the integrity of his own action) to adopt the end of repairing the practice as a whole, or at least of not acting in such a way as to undermine that end. But his aim here is not to make the practice more efficient in producing results that are valued independently of the practice; rather, his aim is to restore the integrity of the practice as a practice, to bring about the conditions under which it would count as the genuinely shared form of activity it purports to

32 Rawls comes close to making this distinction on pp. 32–33 of “Two Concepts of Rules,” where he writes, “...[A]s with any set of rules there is understood a background of circumstances under which it is expected to be applied and which need not—indeed which cannot—be fully stated. Should these circumstances change, then even if there is no rule which provides for the case, it may still be in accordance with the practice that one be released from one’s obligation. But this sort of defense allowed by a practice must not be confused with the general option to weigh each particular case on utilitarian grounds which critics of utilitarianism have thought it necessarily to involve.” This paper can be read as an attempt to interpret this paragraph, and so to shed light on Rawls’s later notion of nonideal conditions.
be. Thus, the existence of a superficial similarity between the stand-
points of the social engineer and the reformer helps to explain why
action under nonideal conditions can appear to be governed by conse-
quentialist reasoning; but further examination of those standpoints
reveals this appearance to be misleading.\footnote{As Tony Laden has pointed out to me, this analysis shows why it is good to build
mechanisms of reform into the structure of practices. Checks and balances which
allow authorities at higher levels to repair damage done at lower levels (for example,
judicial review, constitutional amendment, and so forth) allow participants to change
the rules while still playing by them. In this way such mechanisms narrow the gap
between reformers and social engineers.}

An additional virtue of my account is that it avoids glib moralism,
by showing how a moral dilemma can arise through no cognitive or
volitional failure on the part of the agent. The predicament I have
described is one that retains its dilemmatic character even when the
agent is fully aware of its source and structure. And it does so because
the situation is, at least in one sense, one in which there is no fully
justified course of action to be taken. The agent has to perform a
public action, because to do anything less would amount to failure
to act in his public role, a role in terms of which he values himself
and his conduct. But the situation is such that none of the alternatives
straightforwardly meet that standard. Each of them fails, either for-
mally or substantively, to be the public action it ought to be. In this
sense, there is no right thing to do. I take it this way of looking at
the situation provides grounds for the rather puzzling, retrospective
feeling which Nagel calls “agent-regret.”\footnote{Nagel coins this term in “Moral
Luck,” in his collection Mortal Questions, pp. 24–38, see p. 37.} Agent-regret is the feeling
that you have violated a standard that somehow both applies to you
under the circumstances and is, given the circumstances, impossible
to live up to. On this account, your role as a participant sets a standard
of conduct that applies to you whether or not any of the alternatives
actually open to you meet that standard.

My account is at bottom, however, ambivalent about agent-regret.
For there is another way of looking at the situation in which there is
a right thing for you to do—namely, to do the best you can to give
your action a public character, by choosing the least defective from
a set of defective moves. That might turn out to be an action in
conformity with the letter of the law, or it might not. But as long as
you choose the surrogate that most closely approximates the govern-
ing ideal, you will have done the best anyone in your position could
have done. This way of looking at the situation provides grounds for
thinking that you are free of wrongdoing in any straightforward sense.
On the account I have offered, this fundamental ambivalence about agent-regret is rooted in the fact that the reformer’s unlawful actions can be regarded as differing from lawful actions either in kind or in degree. They differ in kind from lawful actions insofar as they are actions undertaken independently of the practice rules, and hence on the authorization of nothing other than a unilateral will. On this way of looking at such actions, they are simply lacking in public character, and in this respect are no different from the actions of an outlaw. From this perspective, agent-regret may be warranted. But insofar as such actions are regarded as surrogates for lawful actions, undertaken out of your conscientious sense of what the spirit of the law requires in this unusual circumstance, then such actions may be regarded as approximations of lawful conduct, so that the difference between the former and the latter is one of degree. From this perspective agent-regret is not warranted.

Is this ambivalence a virtue or a vice of my account? One reason for thinking it is a virtue is that our ordinary reactive attitudes display the same kind of ambivalence. When an agent has chosen the lesser of evil courses of action open to him under conditions of corruption, we tend to think, on the one hand, that it is inappropriate to hold him responsible for any wrongdoing. Yet we also tend to think that it is at least not inappropriate for the agent, reflecting on his own conduct, to feel that he has somehow compromised his integrity, or that he has been left with “dirty hands.” A theory that unambivalently endorsed or rejected agent-regret would have trouble accounting for one or the other of these intuitions. I believe my account provides a basis for both, without falling into incoherence.35

III.2. Implications for Kantian nonideal theory. I have explained the implications of the view for the problem of moral dilemmas, at least insofar as such dilemmas can arise within the context of social practices. But my primary and longer-term interest is in relating the view I have laid out to Kant’s moral theory in such a way as to show how the scope of Kant’s purism might be limited. This obviously requires a separate article, but the basis of a strategy for doing this has, I think, been put in place. What I believe I have shown is that rules that bind categorically are not for that reason incorruptible, at least when their categorical status depends upon their serving a constitutive function.

35 Given what I have said, this implies that the agent can take the “difference-in-kind” view of his unlawful action, whereas the observer ought to take the “difference-in-degree” view. My account would be strengthened if I could show why it makes sense for the points of view to line up in this way, but at present I do not have such an explanation.
Practice rules can serve as models of categorical imperatives, because they apply unconditionally to all participants. If you are in the role of a participant, then you must play, because that is the only way to act as a participant. And the only way to play is to play by the rules, because the rules are (at least ideally) constitutive of participation.36

Now, practice rules are only models of categorical imperatives, because normally we have the option of abandoning any conventionally-established practice. But what if there were a non-optional practice, a role we had to occupy on pain of not being agents at all, and a set of rules we had to follow on pain of not doing anything at all? If abandoning agency is not a human possibility, then those rules would in the fullest sense be categorical imperatives, and we would be bound to adhere to them under all conditions. But if those rules were categorically binding in virtue of their constitutive role—in virtue of their fitness to constitute a shared will and a shared form of activity attributable to that will—then they would by the same token be vulnerable to corruption; in principle, there would be some threshold below which noncompliance could force the letter and the spirit of the moral law itself to come apart. In this unfortunate circumstance, the conscientious agent would be forced to ask whether his regard for the universal form of his maxim had itself been rendered something of a mere formality.

I have not yet argued for this claim, in large part because many of the actions required by the moral law, for example truthtelling, are not obviously dependent upon anything like a practice to be what they ought to be. A separate argument is thus needed to show how the extension of the argument offered here to such cases as that of the murderer-at-the-door is possible. The argument would have to claim that what is required by the moral law, strictly speaking, is not truthtelling per se but some mode of person-to-person, or, in Kant’s language, legislator-to-legislator relation. In this paper I claim only to have shown that within a practice, there is indeed a threshold beyond which others’ wrongdoing can make your compliance count as a form of complicity.

Tamar Schapiro

Stanford University

36 The strategy of accounting for categorical normative force by appeal to constitutive function has been developed to some extent by Korsgaard, especially in her article, “Self-Constitution in the Ethics of Plato and Kant,” Journal of Ethics, 11 (1999): 1–29. As Korsgaard mentions in that paper, a challenge facing any defender of this strategy is to explain how defective action is possible; how an action which violates a constitutive rule can count as failing to be what it ought to be, instead of simply counting as a successful instantiation of something else. I discuss this in relation to practice rules in my “Three Conceptions of Action in Moral Theory.”