Kantian Rigorism and Mitigating Circumstances*

Tamar Schapiro

I. INTRODUCTION

A task facing any moral theory is to account for both the rigidity and the flexibility of moral rules. It is natural to think of moral rules as being both uncompromising across some range of core cases and open to modification or exception at the margins. Utilitarianism and Kantianism face complementary challenges when it comes to justifying this intuition. Utilitarianism threatens too much flexibility; critics argue that the theory would require us to break our promises, tell lies, and condemn the innocent whenever doing so happens to be expedient. Kantianism, by contrast, seems too rigid. When a murderer comes to the door asking the whereabouts of his intended victim, it seems morally myopic to refrain from lying to him out of a sense of duty. In general, utilitarianism appears to be more attractive than Kantianism in cases that are in some sense marginal or exceptional, cases where we are inclined to depart from our ordinary standards in the face of mitigating circumstances. But Kantianism seems more attractive in what we might call core cases, cases where we are inclined to keep our promises or tell the truth “on principle,” independent of any

* I have revised and, I hope, improved this article in response to comments from many of my colleagues, all of whose names I cannot list here. I am particularly grateful to Allen Wood, Debra Satz, Michael Bratman, Robert Taylor, Henry Richardson, Nancy Sherman, Robert Hannah, Tom Hill, Jerry Schneewind, Susan Wolf, Stephen Darwall, Peter Railton, Aaron James, William Bristow, Julie Tannenbaum, Jens Timmerman, James Mahon, Ethan Leib, Meir Dan-Cohen, Patrick Wilson, and several anonymous referees and editors at Ethics for their helpful responses to earlier drafts.


Ethics 117 (October 2006): 32–57
© 2006 by The University of Chicago. All rights reserved. 0014-1704/2006/11701-0003$10.00
implicit or explicit calculation of the benefit to be achieved or the harm to be averted by doing so.

Given that the two theories face complementary problems, there is a striking asymmetry in the amount of sustained philosophical attention accorded to each. Utilitarians, on the one hand, have long been troubled by the apparent excesses of act utilitarianism and have responded by attempting to develop a coherent form of rule utilitarianism, that is, some form of utilitarianism that would be able to capture the uncompromising character of commonsense morality. Kantians, on the other hand, have addressed the problem of rigorism in a relatively piecemeal and haphazard fashion. Perhaps because rigorism only shows up


in marginal cases, it has been treated as if it had marginal significance for the theory.

Typically, defenders of Kant frame the problem as a relatively superficial matter concerning either the interpretation of isolated Kantian texts or the details of the application of the Categorical Imperative procedure. In this article, I argue that the source of the problem is both deeper and more intuitive. The tendency toward rigorism stems from the basic thought that the actions prohibited by categorical moral rules are "wrong in themselves"—wrong in virtue of the features that make them count as the kinds of actions they are. If an action is wrong in itself, then variations in external circumstances cannot make it right. This general thesis, that rightness and wrongness are intrinsic rather than extrinsic to actions, is common to both Kant's moral theory and rational intuitionist forms of deontology. I will therefore call it “the deontological thesis,” construing “deontology” broadly. In what follows I lay out a strategy for holding on to the deontological thesis while avoiding rigorism. This strategy depends on the idea that moral rules constitute the actions they govern and that this constitutive function is corruptible when background conditions are defective. Moreover, I claim that the actions ideally constituted by the moral rules—for example, acts of honesty—are fundamentally forms of interaction; they must be both addressed and received in a certain spirit in order to count as having integrity as the kinds of actions they are. In conclusion I suggest, but do not argue, that this conception of the relation between moral rules and actions fits more naturally with a Kantian rather than a rational intuitionist interpretation of the deontological thesis.

II. DIAGNOSIS: WHAT IS THE SOURCE OF THE WORRY ABOUT RIGORISM?

I mentioned that Kantians have addressed the problem of rigorism in a relatively piecemeal and haphazard fashion. Part of the reason for this is that there is disagreement among Kant's defenders as to whether rigorism is anything more than a pseudoproblem. Much of the scholarly discussion of the issue focuses on Kant's "Supposed Right to Lie" essay, in which Kant appears to argue that nothing could count as mitigating circumstances when the question is whether to adhere to the duty of truth telling. Now as many readers have pointed out, that text is, on the face of it, squarely at odds with passages in the Lectures on Ethics, in which Kant takes a more permissive stance. Whereas in the "Supposed Right to Lie" essay Kant claims that it is wrong to lie to a would-be murderer even for the sake of preventing his intended murder, in the Lectures he claims that it is permissible to lie to a robber merely for the
sake of protecting one’s money. H. J. Paton suggests that the “Supposed Right to Lie” essay can perhaps be disregarded as a manifestation of Kant’s “bad temper” in his old age. Other commentators maintain that something more philosophically interesting is at stake, but beyond this, diagnoses of the problem diverge. Many point out that Kant’s rigoristic texts, as well as the general worry about rigorism, may stem in part from a tendency to conflate two distinct notions of “unconditionality.” The fact that the Categorical Imperative (CI) is unconditional in its authority, they note, implies nothing about whether it is unconditional—in the sense of “exceptionless”—in its content (no ceteris paribus clauses, no “fine print,” etc.). Moral requirements can be inescapable without being simple. But this rather elementary point cannot be the whole story. Fuller explanations tend to fall into one of two categories. One traces worries about rigorism to exegetical mistakes. The gist of these accounts is that by paying closer attention to the specific contexts in which Kant wrote his more rigoristic texts, as well as to the terminology he uses in those texts, we will see that rigorism is either a pseudoproblem, or less of a problem than it appears. The other, rather pervasive approach, diagnoses worries about rigorism as having their source in systematically misguided methods of applying the CI.

I will not try to address the details of these arguments here, though I believe that most of them contain valuable insights. That said, I am skeptical that any address the deepest source of the worry about rigorism. Arguably, the basic worry arises as readily in response to traditional deontological theories as it does in response to Kant’s theory. If that is the case, then the most general source of the problem should be identifiable independently of specifically Kantian features, like the CI procedure or the interpretive background of the “Supposed Right to Lie” essay. My suggestion is that this source is the deontological thesis that the actions required and prohibited by moral rules are right and wrong intrinsically.

This way of framing the problem has the virtue of bringing out an ambiguity in some Kantian discussions of the wrongfulness of deception. Kant’s defenders stress that the theory prohibits deception not because

of the harm it generally causes, nor even because of the (possibly benevolent) aims with which it is undertaken, but because of its nature as a form of manipulation. Barbara Herman writes, “What the rejection of the maxim of deceit signifies is the inappropriateness of principles that exploit the vulnerability of human agents to manipulative control. Such principles fail to respect the integrity and separateness of will that is constitutive of rational agency.”

It is the character of deception as an assault on rational agency that makes it impermissible, even when it is likely to lead to beneficial consequences. Christine Korsgaard takes this idea a step further, suggesting that the manipulative character of deception is intrinsic to its nature as deception. This comes up in her explanation of why the lying promise is ruled out by the Formula of Humanity. In section 2 of the Groundwork, Kant argues that one to whom a lying promise is addressed is treated as a mere means because he “cannot assent to my mode of acting against him” and is “unable to contain the end of this action in himself.”

These phrases are open to a range of interpretations, and Korsgaard’s reading takes the language of incapacity quite seriously. She writes:

What Kant means, I believe, is that the other person cannot assent to my action because he is not in a position to. This is because he is deceived. By the nature of the case, he doesn’t know how I am acting, and you cannot assent to a transaction you do not know is occurring. In the same way, he cannot “contain in himself” the end of the same action because he is not in a position to. He doesn’t know what the real end of the action is, and is therefore not in a position to make it his own—to choose, freely, to contribute to its realization.

The promisee thinks that the proposal is to promote the promiser’s temporary possession of his money, when in fact it is to promote his permanent possession of that money. Korsgaard’s point is that the lying promiser’s deception, qua deception, necessarily keeps the promisee in ignorance of the content of this proposal and, as a result, makes it impossible for anything the promisee does to count, logically, as freely agreeing to it. Deception is in this sense inherently manipulative, and it is on the grounds of its inherent manipulativeness that it is prohibited independent of considerations of expediency. “Lies are therefore wrong

---

in themselves,” Korsgaard concludes, “regardless of whether they are told with good intentions or bad.”

If deception is wrong because of its manipulative character, and if its manipulative character is partly constitutive of its nature as deception, then deception is wrong in itself. But this appeal to the deontological thesis in turn exerts pressure toward rigorism; if lying or, more generally deception, is wrong in itself, then no change in external circumstances could make it right. How do Herman and Korsgaard avoid this problem?

One way for any deontologist to avoid rigorism is to argue that although right and wrong attach to actions “in themselves,” actions so construed are properly picked out by fine-grained, context-sensitive descriptions, rather than general descriptions like “deception.” Ross employed a version of this strategy in his response to worries about conflicting moral obligations, and Kantians tend to appeal to the same strategy when discussing marginal cases like that of the murderer at the door. Indeed when turning to that problematic example, Herman and Korsgaard seem to back away from the claim that “deception” as such has any determinate moral value. Here they argue instead that terms like ‘deception’ underdescribe actions for the purpose of moral assessment. They support this claim by arguing that the specifically Kantian notion of a “maxim,” properly understood, requires such fine-grainedness. Hence, while a lie to the murderer would indeed be an act of deception, its maxim would differ from other maxims involving acts of deception, such as the maxim of lying for the sake of personal advantage. To finish the argument, they both attempt (in different ways) to show that precise application of the CI rules out self-interested deception while permitting or even requiring deception to thwart the murderer.

For my purposes, the important point is the tension between the claim that deception is wrong in itself and the claim that deception is in itself morally indeterminate until further specified. It seems to me that Herman and Korsgaard appeal to both intuitions without explicitly attempting to reconcile them. The first intuition is attractive in core cases, where the aim is to show that something about the nature of deception, namely, its manipulative character, is what makes it one of the most fundamental forms of Kantian wrongdoing. The other intuition is attractive in marginal cases, where the aim is to show that adverse

12. Ibid.
circumstances can make it the case that some acts of deception are right, or at least not wrong.

III. THE STANDARD KANTIAN ARGUMENT AGAINST PATERNALISTIC DECEPTION, AND THREE PROBLEMS WITH IT

In what follows I will pursue both practical and metaphysical sides of this problem in tandem. Is it possible to hold on to the idea that the wrongfulness of deception stems from features that give it its deceptive character, while still allowing that deception can be made right, or at least not wrong, by the existence of mitigating circumstances? Let me start out by trying to state more clearly just how we have to conceive of “the nature” of deception, to the extent that we take Korsgaard’s metaphysical-sounding language seriously. On her view, deception is wrong in itself, or by its very nature, because it is by definition designed to prevent the deceived from choosing the actions and ends he will pursue. Hence deception, as she characterizes it, necessarily has as its aim the exercise of unilateral control over another person’s will. Implicitly, then, Korsgaard is rejecting a narrower account of the “nature” of deception, one according to which deception is simply the intentional telling of a falsehood. And she is also implicitly rejecting a less normatively loaded account, according to which exercising unilateral control over another’s will is a morally neutral act. Indeed it seems that on her view, deception is wrong by its “nature” only because it has been defined as an act of interference with or usurpation of another’s rightful authority to govern himself. This is not a criticism. I am simply pointing out that Korsgaard’s claim that deception is “wrong in itself” need not be taken as a report of a metaphysical discovery. Rather it is an observation about a certain conception of deception, a conception that presupposes a background of normatively structured interpersonal relations.

I’m going to call this conception “deception-as-interference,” the idea being that deception is necessarily designed to be a way of (illegitimately) interfering with another’s autonomy. Now suppose that this is the right way to think of deception, and suppose that we endorse Kant’s view that interference with another’s autonomy cannot be justified by appeal to good consequences. Is deception then ever permissible, on any grounds? What sort of grounds could justify it? One line of response focuses on cases of paternalistic deception, cases where deception is undertaken for the sake of the good of the person deceived. Even though Kant’s theory rules out paternalistic deception in general, it may be possible to construct a Kantian justification of paternalistic deception in certain kinds of marginal cases: those involving, for example, very young children or the mentally disabled. The claim here would be that these conditions play a mitigating role, weakening the prohibition against deception. But why? If the argument is to work,
these conditions have to be conceived as already interfering with autonomy, albeit in a particular way. Children and the mentally disabled cannot simply be thought of as persons who are likely to make bad choices. For the fact that a person is likely to use her freedom unwisely is not, on the Kantian view, sufficient justification for interfering with that freedom. Instead those who are the objects of permissible paternalistic deception have to be thought of as persons who are not in a position to make choices that count as their own. These would be persons whose deliberative capacities had been hijacked by irrational forces as a result of mental illness, or whose capacities were sufficiently undeveloped so as to be describable as still governed by instinct rather than reason. Both of these conditions mitigate the force of the prohibition on deception, because where autonomy is not intact in the first place, there is nothing for deception to interfere with.15

I take it that this sort of argument is one that most Kantians could agree is, at least in outline, in keeping with the most fundamental orientation of the theory. I will therefore call it "the standard Kantian argument for paternalistic deception." As will become clear, I don’t want to disagree with the standard argument as such. But I do think that it has some philosophically significant ambiguities and limitations, and by focusing on them I hope to gain a clearer account of how deception, or any other inherently manipulative type of action, can become permissible (or obligatory) in marginal cases. I will proceed by identifying three problems with the standard argument. The first two of these problems stem from pushing the question of how the standard argument resolves the dilemma about the locus of moral value. The third problem stems from trying to extend the argument to a different type of marginal case.

How does the standard argument resolve the dilemma about the locus of moral value? Suppose you agree that deception is ruled out in core cases in virtue of its nature, which is to be a way of interfering with another’s authority to make his own decisions. Now imagine that you engage in paternalistic deception of a mentally disabled person, on the grounds that where autonomy is not intact, there is nothing for deception to interfere with. Does this act of deception count as an act of permissible deception, or does it count as an act of a different type? The standard argument, as I am conceiving of it, is ambiguous on this point. But if we take seriously the deontological thesis, then it would make sense to fill out the standard argument in the following way. By hypothesis, your deception does not interfere with anyone’s autonomy, nor is it designed to interfere with anyone’s autonomy. So it isn’t de-

ception after all. Instead we should describe it in some narrower, less normatively loaded way, for example, as an act of “intentional falsehood telling.” When an action properly falls under this description, we might say, it is morally “indifferent,” neither right nor wrong in virtue of its nature. On this construal of the standard argument, as we move from core cases to a marginal case involving someone lacking in autonomy, the nature of the action changes. Because the circumstances that change are partially constitutive of the nature of the deception, what had been describable as deception is now properly described as intentional falsehood telling.

Given this picture, there is no such thing as permissible deception. What is permissible is something that looks superficially like deception but is really merely intentional falsehood telling, a type of action that is intrinsically neither right nor wrong. Now to accept this picture is to grasp one horn of a dilemma generated by the question about the locus of moral value. If we start from the view that deception is wrong by its nature, then no change in circumstances external to the nature of a deceptive act could in principle make it right. But by the same token, changes in circumstances internal to the nature of a deceptive act couldn’t make deception right either, because those changes would alter the nature of the act, making it into something other than deception. Perhaps that is not such a bad thing, though. What is wrong with saying that while there is no such thing as permissible deception, there is such a thing as permissible intentional falsehood telling?

The first problem is that it now looks as though in the shift from core to marginal circumstances, we have abandoned the domain governed by principles of deception and honesty and have moved into an entirely new domain in which principles governing intentional truth telling and intentional falsehood telling apply. This counts as a problem because it seems intuitively that there are constraints on intentional falsehood telling even in cases involving those whose capacities for autonomy have been to some extent undermined, and these constraints, although not identical to those governing deception and honesty in core cases, must in some systematic way be related to the principles we follow in core cases. But on the picture I just described, it is hard to see how one domain of principles would inform the other. It looks as though we need some third set of principles to account for systematic connections between the two domains.

The second problem is this: because the standard argument, as I am imagining it, actually denies the possibility of permissible deception, it implicitly denies the need to employ the concept of an "excuse" to capture a special case of justification. Intuitively, there is a difference in the nature of the question we raise when deciding what to do in core cases as contrasted with marginal cases. In core cases we ask, "what
should I do?” or “is this action justified?” In marginal cases we are conscious of the need to make some kind of exception relative to core cases. The question we ask is not simply “what should I do?” or “is this action justified?” but more specifically, “am I justified in departing from (otherwise uncompromising) standard S in light of mitigating circumstance C?” I believe that this difference in the content of the question raised in marginal cases underlies one way of using the notion of “excuse.” Granted, the notion of an excuse is also, and perhaps more commonly, used in connection with retrospective judgments of responsibility. That use of the notion presupposes that the agent has done a wrongful action, though he is not to be held responsible for it in the usual way. The notion of excuse I will be explicating is applied prospectively. What it aims to capture is not the mitigated responsibility of the agent but, rather, the mitigated wrongfulness of the action. Moreover I will not assume any particular view about the relation between mitigated wrongfulness in this sense and mitigated responsibility.

An excuse, on this conception, is a special case of justification, one in which the agent is aware of having to act in what she conceives of as a marginal case. But on the view I described above, marginal cases are not cases in which this consciousness is necessarily operative. Instead, marginal cases are simply different cases, cases in which the types of actions that are under consideration and the rules governing them are different from the types of actions and rules that are typically under consideration. Other things equal, it would be more philosophically satisfying to have an account that made sense of the difference between a justification and an excuse, understood prospectively, and hence captured the moral agent’s own consciousness of having approached a boundary between core and marginal cases.17

16. This usage is not entirely idiosyncratic. See Korsgaard, “Two Arguments against Lying,” 356. But Allen Wood has, quite understandably, suggested that it would be clearer and more accurate if I called my view a “theory of exceptions” rather than a “theory of excuses.” My reason for not doing so is that an exception is not a kind of justification, and the aim of the theory is to lay out a kind of justification.

17. I take it that most versions of rule utilitarianism or rule consequentialism require a certain lack of transparency between the perspective of the agent and that of the moral philosopher. While the philosopher can recognize the boundary between core cases (in which general rules are to be strictly followed) and marginal cases (in which they are to be set aside), it is not clear that the agent herself is in a position to acknowledge that boundary. That is, it is not clear that the agent herself is in a position to see her situation as one involving mitigating circumstances, one in which she recognizes the need to depart from otherwise uncompromising rules in the face of what she takes to be mitigating conditions. But our consciousness, as agents, of having to act at such a boundary is an important feature of our moral experience, and it is arguably only from this consciousness that certain concepts (e.g., “excuse”) acquire their salience. Hence I believe that in order to gain a proper understanding of such concepts, we need a theory that captures the perspective of the agent.
A corollary of this second problem is this. If there is no such thing as permissible deception, and if there is therefore no conceptual difference between a justification of deception and excuse for deception, then arguably there is no conceptual difference between choosing an act of deception as the right thing to do and choosing it as the lesser of evils. Since it is never permissible to choose an act of deception as the right thing to do, it is also never permissible to choose it as the lesser of evils. This is implied by the view above. What gets chosen in a marginal case is a different act that is the right thing to do and so is chosen as the right thing to do.

But intuitively there are conceptual differences between choosing an action as the right thing to do and choosing it as the lesser of evils. In both cases, an agent determines the action as the most choiceworthy among the alternatives, from a moral point of view. But in the case of a choice of the lesser of evils, the agent necessarily regrets the choiceworthiness of the action he sees as most choiceworthy. So, for example, to choose deception as the lesser of evils is both to recognize the choiceworthiness of deception under the circumstances and to regret the fact that circumstances have made deception count as the choiceworthy alternative. The same cannot be said of the case where an agent chooses deception—or even intentional falsehood telling—simply as the right thing to do.

Notice that in the case of the choice of the lesser of evils, the object of regret is not the circumstances as such, nor the agent’s own performance of the action, but the choiceworthiness of the most choiceworthy action. This leads to another difference between choosing an action as the right thing to do and choosing it as the lesser of evils. Whereas circumstances provide the context for every choice of action, it is only when we choose an action as the lesser of evils that we are inclined to say that “circumstances forced us” to choose the action. If there is no distinction between a justification and an excuse, and no distinction between a right action and a lesser of evils, then this is a curious thing to say. For circumstances always play a role in the determination of the

18. Although I mention regret, I do not intend this passage to be an explanation of Nagel and Williams’s notion of agent regret. (See Bernard Williams, “Ethical Consistency,” in his *Problems of the Self: Philosophical Papers, 1956–1972* [New York: Cambridge University Press, 1973], 166–86; Thomas Nagel, “Moral Luck,” in his *Mortal Questions* [Cambridge: Cambridge University Press, 1979], 24–38, 37.) My interest here is in defining a distinction between “the lesser of evils” and “the right action” independently of considerations about how choosing the lesser of evils bears on the agent’s responsibility. Moreover, I do not claim here to capture the structure of every type of moral conflict that might be called a “moral dilemma.” Nor do I claim that every instance of acting in nonideal conditions is one where perplexity is appropriate. In some cases it may be clear which action counts as the lesser of evils, and that of the lie to the murderer may well be one of those.
Schapiro  Kantian Rigorism and Mitigating Circumstances  43

choiceworthiness of an action. In cases where we regard the choice-
worthy action as simply the right thing to do, we do not say that "cir-
cumstances forced us" to choose the action. There is a kind of necessity
at work in all choices of right action, but it is simply the necessity of
obligation. In any set of circumstances, principles (along with, perhaps,
some account of the identity or nature of the agent) will determine
what counts as the most choiceworthy thing to do from the moral point
of view, and if moral obligation is genuinely binding, then this complex
of principles and circumstances will determine what we ought to do
(regardless of what we might happen to want to do). And insofar as we
are responsive to morality, we will in fact do what we ought to do. Yet
when we say "circumstances forced us" to choose an action, we are
making reference to something other than the ordinary force of moral
obligation. I take it that what I have already said about the nature of
the choice of a lesser of evils helps us begin to explain this. If in that
choice we necessarily regret the choiceworthiness of the action that is
determined to be best from a moral point of view, then in saying "cir-
cumstances forced us," we are attributing the deficiency in the choice-
worthy action to some sort of privation in the circumstances rather than
to a privation in the principles we have applied to those circumstances.
This latter feature is something yet to be explained by a plausible ac-
count of the nature of excuse, and I believe that the full account I will
be defending will shed light on it.

To recap: the first problem with the standard argument as I de-
scribed it is that it leaves us wondering what to do in marginal cases,
and this is so because it accounts only for the normative discontinuity
between core and marginal cases, while not obviously accounting for
the normative continuity. The second problem is that it does not dis-
tinguish excuse (in its prospective use) as a special case of justification
and that it cannot make corollary conceptual distinctions, for example,
between the right action and the lesser of evils, and between noncoercive
and coercive sets of circumstances. In addition to these there is a third
problem with the standard argument. It is that the justification of pa-
ternalistic deception does not extend unproblematically to another par-
adigmatic case in which an exception to the prohibition on deception
seems warranted. This is the case of the murderer at the door, which I
will consider in general terms as a case in which defensive deception
appears to be warranted. Defensive deception, as I will be conceiving
of it, is deception undertaken to thwart another’s wrongdoing.19 For the
sake of simplicity I will be considering cases where the person to be

19. This is not to be confused with punitive deception, or punishment generally,
which is a different sort of action. Punishment, as I see it, is essentially a way of holding
another person accountable for his wrongdoing. This is not the aim of defensive deception.
deceived is himself the wrongdoer, as in the case of the murderer at the door, though arguably cases of defensive deception might arise in which the deceived is not also the wrongdoer. Those cases would, I think, raise problems additional to the ones I will be focusing on in this article.

Can the argument used to justify paternalistic deception—or rather, paternalistic intentional falsehood telling—succeed in justifying defensive deception, or defensive intentional falsehood telling? To extend the argument, it looks like we would have to claim that the murderer at the door is not capable of exercising his autonomy, that he is not in a position to make choices that count as his own. This seems the wrong thing to say, especially if we are committed to holding the murderer accountable for his evil conduct, should he carry it out. The murderer is not a child, and even though he may be mentally disabled, to the extent that we are ready to hold him accountable we are committed to thinking of him as having chosen his evil conduct. If we were to deceive him, moreover, we would certainly have the aim of interfering with his ability to act on his own choices. So it looks as though we need an entirely different sort of argument to justify defensive deception, and it is unclear what form that argument could take.

A natural thing to say is that we ought to deceive the murderer so as not to become complicit in his evil plan. This makes intuitive sense, but it sidesteps the central philosophical problem. Any notion of complicity has to presuppose some notion of moral responsibility. We do not become complicit in wrongdoing simply by being links in the causal chains that eventuate in wrongdoing. We are complicit only if it makes sense to hold us accountable for the contribution we make to the wrongdoing. Now both Kant’s theory and traditional deontological conceptions of morality put limits on an agent’s responsibility to prevent the wrongdoing of others. The fact that you act wrongfully in response to my right action does not in itself make me responsible for, or complicit in, your misconduct. Moreover, even though I do have a responsibility to refrain from manipulation, this does not derive from a general duty to minimize acts of manipulation, a duty that would permit me to manipulate whenever doing so is necessary to lessen the overall number of manipulative acts. So I am not made complicit simply by the fact that I have not acted so as to minimize wrongdoing. Nor am I made complicit by the fact that I have not acted so as maximally to promote the best moral state of affairs. Even though I have a duty to contribute to the promotion of peace and justice, that duty does not include a permission to promote those ends by manipulative or unlawful means. In sum, if the justification for defensive deception is that honesty would make us complicit in wrongdoing, an explanation needs to be given for why
honesty in this instance, rather than in others, makes us complicit in another’s freely chosen wrongdoing.

So the third problem is this: given that the wrongdoer against whom we are inclined to undertake defensive deception is indeed in a position to exercise his autonomy, it is hard to see how the standard argument can be extended to inform our justification of defensive deception. Moreover the justification cannot be that we must lie to avoid complicity, unless further argument is given as to why, in this case, our otherwise limited responsibility does not end where the murderer’s self-determination begins.

IV. PATERNALISTIC DECEPTION REVISITED

In the remainder of this article, I will propose a revision of the standard argument that I believe serves as the basis for a theory of excuses consistent with the deontological thesis. This theory is intuitively plausible in its own right, and it will have resources that allow it to solve all three of the problems outlined above. The first step is to frame the standard argument explicitly as a response to a request for an excuse, in the prospective sense I referred to above. Instead of seeing the argument as an answer to the question “what makes deception permissible?” we should see it as answering the question “what makes it permissible to depart from a purportedly categorically binding standard of honesty in our relations with others?” The question is why it is permissible to choose deception over honesty, given the purportedly uncompromising nature of the Kantian demand to be honest in core cases.

Now the answer, of course, cannot be that departures from a standard of honesty are warranted when honesty is likely to result in bad consequences. This is the point of maintaining that children and the mentally impaired are not to be conceived simply as persons who are likely to make bad choices. But what is the alternative? The alternative, I want to suggest, is that departures from honesty are warranted when engaging in honesty is impossible. This sounds strange, of course, because there is a narrow conception of honesty according to which it is almost always possible to be honest. If honesty is simply intentional truth telling, then it is certainly possible to be honest with a very young child or someone who is severely mentally impaired. The question, then, is whether this narrow conception of honesty is the relevant one. Now

20. I take this to be a particular way of elaborating Korsgaard’s comment on the standard argument for paternalistic deception: “In order to annul the procedural requirement that each person decide for herself, we must show that the procedure could not in any case be carried out” (“Two Arguments against Lying,” 349). However unlike Korsgaard, what I am identifying as “the procedure” is honesty, regarded as a composite of letter and spirit. The spirit of honesty does, of course, have something to do with self-government. In Sec. V I will have more to say about this.
recall that the standard argument relied on a conception of deception as a way of interfering with another’s autonomy. As such, that argument implicitly relied on a conception of honesty as a way of acknowledging or respecting (by not interfering with) another’s autonomy. Call this “honesty-as-noninterference.” If it is the nature of honesty to be an act of noninterference, is it possible to be honest with someone who lacks autonomy in the first place?

The question is the flip side of the question I pushed earlier, whether deception of the nonautonomous counts as deception or as some other kind of action. The first two problems I laid out stemmed directly from filling out the standard argument in a particular way, by imagining that the answer was yes, it is impossible to count as deceiving the nonautonomous. So the corollary answer would be that yes, it is impossible to be honest with the nonautonomous. Both honesty and deception, we might imagine, are defined with reference to interactions with a person assumed to be autonomous, while in the case of the nonautonomous the relevant action-types are intentional truth telling and intentional falsehood telling, and these are governed by different principles.

Since this way of thinking led to the first two problems, it makes sense to ask whether there is an alternative to it. But the task is to find an alternative that does justice to the intuition that the inherently manipulative nature of deception is what makes it wrong, while the inherently nonmanipulative nature of honesty is what makes it right. The solution is to see the narrower conceptions of honesty and deception, those of intentional truth telling and intentional falsehood telling, not as different types of actions, but as deficient forms of honesty and deception, respectively. The claim is that in cases where the person to be deceived lacks autonomy, honesty is reduced to mere truth telling, and deception is, in a sense, reduced to mere falsehood telling. And the idea behind that claim is this: where those with whom we interact are lacking in autonomy, the background constitutive context that gives honesty and deception their proper nature is rendered deficient. As such, the actions themselves are corrupted, without thereby being changed into different types of actions subject to different internal norms. One helpful way to phrase this is in terms of a contrast between the “letter” and the “spirit” of each type of act. The nature of an act of honesty, we might say, is to be a composite of letter and spirit, in the sense that it is to be an act of intentional truth telling undertaken as a way of acknowledging (by not interfering with) another’s authority as a self-governor. The act is honest in letter insofar as it is an act of intentional truth telling. It is honest in spirit insofar as it is undertaken as a way of acknowledging another’s right to govern himself. This way of thinking about intrinsically right actions, as composites of letter and
spirit, can explain our intuitions about right actions done for wrong motives. It makes sense to say, for example, that an act of honesty undertaken for purely strategic reasons is honest in letter only, and not in spirit. What certain marginal cases reveal is that there is another way in which an act of honesty can be rendered deficient, and that is when the person to be deceived is lacking in autonomy. For in that case an act of honesty can be honest in letter only, in the sense that it is merely an act of intentional truth telling. It cannot be undertaken as a way of acknowledging another’s authority as a self-governor, because the person is, by hypothesis, not in a position to govern himself. As such it fails to be an act of honesty in the full sense.21

The point can be illustrated with a political analogy: to view someone as lacking the capacity for autonomy due to impairment or underdevelopment is to view him as being governed by alien forces, whether they be powerful unconscious emotions or natural but uncultivated instincts. Such persons are, from a practical point of view, analogous to nations under the occupation of a foreign and arbitrary power. Their representative institutions have been, at least in some respects, usurped and pressed into the service of the alien power. The upshot is that it is impossible to engage in forms of interaction the character and value of which presuppose the integrity of those representative institutions. It might be possible to follow the letter of diplomatic law, but by doing so one would, at worst, be betraying the sovereign will that law is designed to acknowledge. And even at best one would be engaging in somewhat empty formality, a procedure lacking in substance. Similarly, I want to argue, the propriety of honesty can in principle come to be hollow when those to whom the truth is told are already held hostage to nonrational forces. Honesty is in these cases reduced to mere truth telling. It becomes honesty in letter only, but not in its interpersonal spirit.

This is not to say that deception is always warranted whenever the one to be deceived is incapable of autonomy. The point is that under these circumstances, both honesty and deception become, in a sense, impossible, because both types of action are defined and gain their moral significance relative to a background in which the parties to the interaction are conceived as capable of self-government. What, then, is a conscientious person to do in such cases? Since the problem is that the letter and the spirit of honesty come apart, the solution is to find what I’ll call a “surrogate” for honesty, to improvise a form of conduct that would come as close as possible to conforming to the ideal of human

21. There is an analogy between this type of failure and the “misfiring” of a speech act due to a privation of background “felicity conditions.” See J. L. Austin, How to Do Things with Words (Cambridge, MA: Harvard University Press, 1962), 16.
relations that animates acts of honesty in the first place. This may require violating the “letter” of honesty, or it may not. If it does, then deception is permissible, though subject to a substantial constraint. The constraint is that the deception be undertaken in the spirit of honesty. That means deception has to be undertaken in such a way as to avoid interference with autonomy to the extent that it does exist and to foster or at least not preclude the development of autonomy where it does not yet exist. At the same time the deception must at the very least be designed to promote the good of the person being deceived, for that is a way of respecting the person as a source of value, even if not a self-governing one. That said, it is also possible that in many cases the best surrogate for honesty is honesty in letter only, truth telling undertaken in an “aspirational” mode, as an attempt to model and foster the kind of equal relationship that acts of honesty properly require as their constitutive background.22 There is, no doubt, a virtue involved here, a capacity to find the mean between the vices of purism and pragmatism when faced with the problem of how to cope with conditions inhospitable to moral ideals. I’ll leave the project of defining that virtue for another day.

This way of understanding the standard argument shows how some changes in circumstance, while not altering the nature of honest actions (changing them into a different type of action which would be subject to different internal norms), can nevertheless corrupt them by detaching them from the constitutive background against which they acquire their animating spirit. As a result honest actions are reduced to acts of mere truth telling. The conclusion is this: if deception is wrong in itself yet sometimes permissible, this is because honesty, though intrinsically right, is nevertheless corruptible. That said, even when deception is permissible in letter, it is not for that reason permissible in spirit. At best deception is permissible only under certain constraints, as a way of approximating the spirit of honesty. Moreover, the fact that deception can be undertaken in the spirit of honesty shows that it, too, is “corruptible,” in the sense that it can be detached from the interpersonal background that gives it its proper significance and value. In this respect the corruptibility of deception turns out to be just as essential to the excusability of deception as is the corruptibility of honesty.

This revision of the standard argument helps to solve the first two problems I identified earlier. First, it shows how constraints that apply in core cases are still relevant in marginal cases. As I already noted, when faced with corrupting circumstances, we appeal to a conception of the action’s integrity, in this case to a conception of honesty as a

22. I thank Seana Shiffrin for suggesting use of the term ‘aspirational’ in this context.
composite of both letter and spirit, to guide our judgment in finding a surrogate for the action in marginal cases. Second, this version of the standard argument carves out conceptual space for a distinction between a justification and an excuse. An action is excused, in the sense that its intrinsic wrongfulness is mitigated, when it is chosen as a surrogate for another action as a way of coping with the corruption of that action. As such, there is conceptual room to draw a distinction between choosing an action as the right action and choosing it as the lesser of evils. For a surrogate action, even when it is the most choiceworthy under the circumstances, is always to some extent deficient; it necessarily falls short of being a realization of the relevant ideal in both letter and spirit. Hence, for example, when deception in the spirit of honesty turns out to be the most choiceworthy course of action, it still makes sense for the agent to regret the choiceworthiness of this way of realizing the spirit of honesty. Finally, this account helps to shed light on why we say, in marginal cases, that circumstances forced us to choose the action that was most choiceworthy under the circumstances. To say this, as I mentioned earlier, is to attribute the deficiency in the choiceworthy action to the circumstances rather than to our principles. This makes sense since the view depends on seeing marginal circumstances not simply as different conditions but as deficient conditions, conditions that undermine the integrity of actions that ought properly to be right, rather than conditions that simply make it the case that different types of actions are right.

V. HOW THE REVISED ARGUMENT CAN BE EXTENDED TO DEFENSIVE DECEPTION

What I have shown so far is that it is possible to take seriously the idea that certain actions are wrong “in themselves” while still accounting for their permissibility by appeal to the existence of corrupting background conditions. I have also argued that this allows us to account for permissibility in the right way, in a way that captures some of the distinctive features of moral choice under marginal circumstances. Now it is time to turn to the third problem with the standard argument: the problem of how to extend it to cases of defensive deception. If the murderer at the door is, by hypothesis, not lacking in autonomy, how can deception of him be excused? Do we need an entirely different argument to deal with this kind of case?

I believe that the standard argument can indeed be extended to cover defensive deception. But the extension requires us to revisit our conception of the nature of deception, which I earlier referred to as “deception-as-interference.” Recall that on that view, it is the nature of

23. But see n. 18.
deception to be a way of interfering with another’s autonomy, and cor-
relatively, it is the nature of honesty to be a way of acknowledging (by
not interfering with) that autonomy. In order to extend the argument
to defensive deception, we need to regard this conception as being, not
incorrect, but incomplete. Deception-as-interference focuses on the fact
that my act of deception prevents you from making your own choices.
But in preventing you from making your own choices, I prevent us from
making choices that count as ours, choices that stem from a shared
practical standpoint. Deception, by its nature, prevents us from sharing
such a standpoint. This line of thought suggests a fuller conception of
deception, one according to which deception, insofar as it interferes
with another’s autonomy, amounts to a refusal to reciprocate within a
scheme of shared thought and action. I’ll call this conception “decep-
tion-as-refusal.” Like deception-as-interference, deception-as-refusal pre-
supposes that deception takes place against a background of normatively
structured relations. But deception-as-refusal presupposes a more robust
form of reciprocity among autonomous individuals than that suggested
by the “interference” picture. Whereas deception-as-interference pre-
supposes a conception of the moral community as an aggregate of self-
governors bound by obligations of noninterference, deception-as-refusal
presupposes a conception of the moral community as a realm of joint
participation in shared activity. In Kantian terms, it presupposes that
the moral community is a Kingdom of Ends, a realm in which each
autonomous member participates along with every other in the activity
of self-legislation.

Now recall the problem: the murderer at the door is not lacking
in autonomy; he is in a position to make his own choices and to be
held accountable for them. So it looks as though it is indeed possible
to be honest with him in both letter and spirit. That is, it is possible
to tell him the truth intentionally as a way of refraining from inter-
ference with his self-government. The standard argument used to ex-
cuse paternalistic deception in the case of the mentally disabled cannot
excuse defensive deception in the case of the murderer at the door.
But suppose that what underlies the standard argument is a more
fundamental conception of deception-as-refusal. Suppose, in other
words, that the point of refraining from interference with another’s
capacity to think for himself is that such restraint is necessary if you
are to share a common deliberative standpoint with him. If this is the
case, then we have not yet fully articulated the spirit of honesty. The
spirit of honesty, according to this line of thought, is not simply to
refrain from interfering with another’s self-governing activity, or even
to express acknowledgment of the other as a self-governor, but rather
to live up to an (implied) demand to think and act from a shared
standpoint. To see honesty this way is to see it as a form of address
issued by one colegislator to another. So conceived, honesty is a mode of interaction that takes its character from the interpersonal relation from which it issues, a relation characterized by reciprocity as well as autonomy. If this is assumed, then the story we have told about paternalistic deception still holds; when the other is not in fact fit to govern himself, then the character of honesty as a form of address to a colegislator is ill-established. But now we can go on to ask whether lack of autonomy is the only way in which the background presupposed by honesty can be corrupted.

Just as in the case of paternalistic deception, we cannot simply say that you are entitled to deceive the murderer because if you were honest with him, he would be likely to make a very bad choice. And yet if there is anything about the murderer’s condition or orientation that could serve to mitigate the prohibition on deception, it would have to have something to do with the fact that the murderer is a bad guy, that he has adopted and plans to pursue a blatantly immoral (some would say “evil”) end. This makes Kant’s defenders nervous, because there seems to be no way to acknowledge the latter intuition without tacitly appealing to the principle that we are allowed to deceive others whenever we judge their choices to be bad, or bad to a certain degree. In Korsgaard’s discussion of the lie to the murderer, she is clearly alive to this worry. There she maintains, somewhat counterintuitively, that the excuse for the lie does not depend in any way upon the evil content of the murderer’s intentions, but rather on the fact, which she takes to be implicit in the case, that the murderer is trying to deceive you about the content of his intentions. But what if the murderer wasn’t trying to deceive you about his intentions? Would it then be wrong to try to deceive him? Surprisingly, Korsgaard accepts this conclusion. She explicitly states that if the murderer were straightforward with you about his evil intention, you would no longer have an excuse to lie (assuming you cannot avoid answering yes or no). She then tries to minimize the cost of bullet biting by suggesting that such a case would never arise. But it certainly could arise. Suppose that a Nazi came to the door under the mistaken impression that you are a good Aryan who believes strongly in the Nazi cause and that you are harboring a

24. For help in articulating this point, I am indebted to Stephen Darwall’s work on the second-person standpoint. See his The Second Person Standpoint: Morality and Accountability (Cambridge, MA: Harvard University Press, 2006). In Darwall’s terms, I am claiming that “deception” and “honesty” are second-person concepts. More generally, I think my view implies that to the extent that intrinsic value can be said to attach to actions, those actions have to be describable in terms of second-person concepts and their associated second-personal relations.


26. Ibid., 137, n. 3.
Jew simply out of ignorance of his identity. This Nazi would have every reason to be completely up-front about his intentions on the assumption that you sympathize with them. On Korsgaard’s view, his straightforwardness would make it impermissible for you to lie to him.

What is notable, for my purposes, is the philosophical motivation that drives Korsgaard to this position. What leads her in this direction is her legitimate concern to avoid the slippery slope that seems to follow from claiming that the substantive moral value of the murderer’s intentions makes a difference to how you should treat him. This would conflict with the idea that we owe others honesty simply in virtue of their humanity, the rational capacity that gives them their status as persons. But is there a more intuitively satisfying way to avoid the slippery slope? A better account, I think, would go like this. The content of the murderer’s intention is relevant to the excuse, because it reveals something about the orientation of the murderer’s will that is relevant to whether or not you can be honest with him in both letter and spirit.

More specifically, his adoption of an end that is blatantly at odds with the ideal of reciprocity in a Kingdom of Ends shows that he is not in a position to take up your honesty in the spirit in which it would be offered, namely, a spirit of reciprocity. His choice of that end makes it appropriate for you to regard him as having refused to participate in the shared activity of which honesty is a part, the activity from which honesty derives its moral value. This is so despite the fact that on the surface, the murderer is treating you with respect. The fact that he intends to murder an innocent person shows that whatever game he is playing, it is not the colegislation game, even though in certain respects it may simulate the colegislation game. Now if honesty is conceived as a form of interaction rather than simply a form of action, then the murderer’s lack of reciprocity will have a bearing on your own capacity to be honest with him. If the murderer has refused to play the colegislation game, then whatever demand he is issuing, it is not the demand to share in thought and action, the demand to which honesty is properly addressed. Hence it is impossible for you to be honest with him in the spirit proper to honesty.

This excuse for defensive deception is consistent with the one I defended in the case of paternalistic deception. Deception (in letter, though not in spirit) is justified when honesty in letter and spirit is impossible, and that is the case when the other is not in a position to engage in the form of interaction of which honesty is a part. The phrase ‘not in a position to’, as I use it here, can mean either “can’t” or “won’t.” In paternalism cases, others simply can’t participate due to conditions of disease or immaturity. In defense cases, they simply won’t participate; it is through their own choices that they are “not in a position” to engage.
Both represent forms of recalcitrance with which the conscientious agent has to cope.27

Although I do not claim to have fully defended a conception of deception-as-refusal, or a conception of the moral community as a sphere of shared activity, I have tried to show how such conceptions, taken together, could be invoked to give the right kind of excuse for deceiving evildoers—an excuse that allows us to hold onto the deontological intuition that deception is wrong in itself. But I have not addressed the slippery slope worry directly. How evil does the end have to be for us to regard its chooser as having refused to reciprocate in the relevant sense? If the choice of any bad end is sufficient to mitigate our obligations, then we should be entitled to engage in defensive deception, or defensive manipulation more generally, whenever those with whom we are dealing are ready to do wrong of any kind. Are we in general to decide how to treat others on the basis of an evaluation of the goodness of their ends?

What we need is a principled way of marking a threshold between the cases in which others’ wrongdoing has no bearing on their status in relation to us and cases in which it does. Notice that an analogous problem arises in the paternalism cases. In order to excuse paternalism, we have to argue not simply that the other is likely to make a bad choice but, rather, that he is not in a position to make choices that count as his own. Instead of seeing him as one who governs himself badly, we have to see him as someone who can’t govern himself. This suggests an analogous way of marking the relevant threshold in defensive cases. Instead of seeing the wrongdoer as one who plays the colegislation game badly, we have to see him as someone who won’t play. The threshold is between imperfect compliance and noncompliance.

One way to get an intuitive handle on this distinction is with reference to personal relationships. Relationships like friendship and marriage are structured by constitutive rules. These rules are expressive of the ideals in light of which we see friendship and marriage as valuable, and it is by complying with these rules, more and less perfectly, that we make our friendships and our marriages more and less perfect reali-

27. I take this distinction to be the deep source of the two parts of nonideal theory. Rawls identifies these as “natural and historical limitations” and “partial compliance,” but apart from maintaining that these are conditions that interfere with the realization of justice, he does not give a deep account of the principle by which they come to count as interfering conditions. See John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), esp. sec. 39, and *The Law of Peoples* (Cambridge, MA: Harvard University Press, 2001), 89–120. The position I defend here is significantly influenced by Rawls. For an earlier, more limited account of my view as it applies to institutional contexts, see my “Compliance, Complicity, and the Nature of Nonideal Conditions,” *Journal of Philosophy* 100 (2003): 329–55.
zations of their kind. But such compliance is often (if not always) imperfect. Friends don’t always listen as attentively as they should, and they aren’t always as thoughtful or generous as friends ideally should be. These are among what I will refer to as the “offenses” that tend to render friendships imperfect. Now an offense, I want to claim, is different from another type of violation, which I will call a “betrayal.” An offense issues from the standpoint of one whose basic commitment to the relationship is not in question. As such it has a bearing on the degree of perfection of the relationship, but it does not undermine the relationship’s basic integrity. A betrayal, by contrast, issues from the perspective of one who is legitimately subject to the demands of the relationship, but whose fundamental commitment to the relationship is in question. As such, betrayals throw the basic character of the relationship into question. To see a friend’s violation as a betrayal is to see it as raising the question whether you and he are really engaged in a friendship, or whether you are merely going through the motions of friendship. It is to ask whether the friendship still exists in substance, or whether it has become a friendship in name only.

The excuse for defensive deception can avoid the slippery slope objection by appeal to the distinction between an offense and a betrayal, applied in this case to the basic moral relationship. If the content of the murderer’s end is relevant to our assessment of his standing relative to us, it is because through the choice of that end he has revealed a basic lack of commitment to the moral relationship. This lack of commitment mitigates our obligations to him, because it undermines the integrity of the moral relationship that is a condition of the integrity of the forms of interaction, like honesty, to which we ideally attach moral value. At a conceptual level, then, the distinction between a betrayal and a mere offense can serve to distinguish between wrongdoers whose wrongdoing does not excuse manipulation on our part, and those whose wrongdoing puts us in a position to engage in constrained forms of manipulation.

I say “constrained” forms of manipulation, because even if our moral obligations to the murderer have been attenuated, they have not been eliminated. The murderer has not, through his betrayal, put himself “beyond the pale” of the moral community. By betraying the moral relationship, he has not thereby exited it. Betrayal and exit are different types of acts. Betrayal is necessarily illegitimate, whereas there may be legitimate ways to exit a relationship. Moreover, unlike personal relationships such as friendship and marriage, participation in the moral community is not optional. There is no such thing as legitimate exit from the moral community; regardless of whether or not you are com-
mitted to the colegislation game, you are still required to play. Hence the murderer has not, through his betrayal, liberated himself from the demands of morality. And by the same token, he has not liberated us from all moral demands with regard to him. He has not somehow turned himself into a nonperson, an animal, or a barbarian. He is neither a nonplayer (in which case we would have no obligations to him), nor an imperfectly compliant player (in which case we would have fully reciprocal obligations to him). He is a player who is refusing to play. He has, through an act of his own volition, put himself in a condition of noncompliance.

If this is right, then it may indeed be impossible to be honest with the murderer in both letter and spirit. If deception is the refusal of reciprocity, then honesty, as I said earlier, is the fulfillment of an implicit demand to deliberate together from a common standpoint. If the murderer has to be regarded as having withdrawn his commitment to reciprocity, then we cannot see him as having demanded that you share a common deliberative standpoint with him, qua colegislator (rather than, say, qua Aryan). He has gone through the motions of such a demand, but the orientation of his will is at odds with the spirit of the transaction. As such you cannot intentionally tell him the truth as a way of fulfilling an implicit demand, issuing from his standpoint, to think together with him. Indeed by doing so you will merely be going through the motions of honesty while allowing yourself to become a participant in his betrayal. It becomes honesty in letter only, at best a merely aspirational use of honesty and at worst a form of complicity with the unilateral spirit of his actions.

Again, under these conditions the best you can do is to improvise, to try to find a surrogate form of action that will come as close as possible to realizing the ideal of human relations that animates your normative conception of honesty. In this case an aspirational use of honesty may be so grossly at odds with that idea that it becomes obligatory to deceive the murderer. But again, even if that is the case, it is imperative to engage in deception from the spirit of honesty. This means engaging in deception in a defensive spirit, not as a way of refusing reciprocity, but as a way of blocking another’s attempt to make it impossible for you to act in the spirit of reciprocity. This would imply, in general, that you are not allowed to engage in gratuitous manipulation of the murderer. Gratuitous manipulation would be manipulation that would not

28. By saying this I don’t mean to be committing myself to the view that moral reasons are “external” in Williams’s sense (“Internal and External Reasons,” in his Moral Luck: Philosophical Papers, 1973–1980 [Cambridge: Cambridge University Press, 1981], 101–13). I leave open the possibility that the inescapability of morality may be rooted in the inescapability of agency.
be necessary either to mitigate the ill effects of his betrayal or to make it possible for him to recommit to the moral community. So, for example, even if you are entitled to lie to him about the whereabouts of his victim, and even if you are allowed to detain him coercively until the police arrive, you are not allowed to torture him for the sake of indulging your indignation.

VI. CONCLUSION

In this article I have tried to illustrate a distinctive approach to the problem of rigorism by looking closely at the prohibition on deception. I have argued that this approach to this problem accounts for the deontological intuition that deception is wrong “in itself,” while still allowing for the possibility of permissible deception in marginal cases. The claim is that both honesty and deception are, in the first instance, forms of interaction and that they are defined against a normatively structured background relationship of reciprocity between autonomous agents. As such the integrity and moral value of honesty depend upon its being both addressed and received in the right spirit, as a way of realizing and upholding that relationship. For this to be the case, those with whom we undertake to be honest have to be both autonomous and committed to reciprocity, at least in a minimal sense. These are “ideal” conditions that make it possible for us to be honest with our interlocutors. Where ideal conditions are not fulfilled—where those with whom we interact either cannot or will not participate in the background relationship in the requisite, minimal way—the only form of honesty open to us is degenerate, and it is on this ground that constrained forms of deception can become permissible.²⁹

²⁹. Am I claiming that this—the appeal to nonideal conditions—is the only ground on which constrained forms of deception can become permissible? Intuitively there are problematic cases in which exceptional deception seems justified, even though the autonomy and reciprocity conditions are met. Consider two of these cases, suggested to me by an anonymous editor of *Ethics*: (1) imagine that Jane wants to rush into a burning building to recover something of great sentimental or aesthetic value (letters, a painting). Or imagine that Joe wants to rush into the same building to recover the last existing pills that can mitigate his terrible and disabling back pain. Now imagine that Tony wants to bungee jump from a particularly high bridge using a particularly dangerous form of cord in order to get a great thrill. Finally, imagine that Robert wants to go out under enemy fire to recover a comrade’s body. Assume that the risk of death and injury to Jane, Joe, Tony, and Robert is the same and that each has an adequate understanding of those risks. Most of us would think it more appropriate to interfere with Tony’s decision (even by means of deception, if necessary) than with Jane’s, Joe’s, or Robert’s. This intuition does not seem to be justified by the thought that Tony is not autonomous, while the others are. Rather, it seems to be justified by our sense of the relative importance of the values at stake in Tony’s decision. (2) Suppose that you lie to Jack about Jill’s whereabouts because you know Jack has a tendency to talk too much and you want to protect Jill from a lurking stalker. You are not acting paternalistically to advance Jack’s good, and you don’t deny
Notice that on the view I am proposing, commonsense moral rules such as “be honest” are to be regarded neither as absolute commands to be followed blindly, nor as mere rules of thumb to be valued for their efficacy. Instead they function as constitutive rules of the moral relationship. What I have tried to explain is how such rules can be both rigid in core cases and flexible in marginal cases. The rigidity comes from the fact that they are constitutive rather than instrumental rules; their role is to construct actions-as-realizations-of-ideals rather than to identify preexisting actions that serve as means of promoting goals. The flexibility comes from the fact that unlike ideals themselves, actions-as-realizations-of-ideals are vulnerable to corruption.

The larger implication of this, I think, is that the deontological thesis ought to be interpreted in a way that is arguably closer to Kantianism than to rational intuitionism. According to rational intuitionism, the actions governed by moral rules have intrinsic value, but they are constituted independently of those rules. The function of the rules is to detect the actions that bear intrinsic value, and the point of acting on those rules is, in some (usually unspecified) sense, to “conform” to that value. By contrast, on the view I have just laid out, moral rules constitute the actions they govern, so as to make them fit to fulfill a function in which their intrinsic value consists. While I cannot here defend the claim that this is a more Kantian construal of the deontological thesis, that suggestion does not seem to me to be implausible.

---

that Jack is an adult with all the usual adult capacities. Nor do you think that Jack is improperly motivated. He just has poor judgment about when to talk. My relationship with Jack is in no way “corrupted,” and yet if the potential negative consequences to Jill were great enough, your deception of Jack would, intuitively, be justified. These examples raise the general question whether all “catastrophe” clauses are to be explained by my account. Am I claiming that all cases in which exceptions seem to be justified on grounds that they are necessary to avoid particularly horrible consequences are in fact to be justified by appeal to the kind of nonconsequentialist reasoning I have been laying out here? Or am I claiming merely that at least a subset of such cases can be explained in this way? What I have said in this article only supports the latter, weaker claim. But then I do owe the consequentialist an account of what is going on in the remaining cases. One strategy would be to argue that aside from the autonomy and reciprocity conditions, there are additional “ideal” conditions, and that it is their failure to hold that explains our readiness to make exceptions in the remaining cases. The alternative strategy would be to claim that the nonideal approach is only part of the nonconsequentialist approach to exceptions and that other cases can be explained by appeal to some other nonconsequentialist strategy or strategies. In this article I do not commit myself to either of these strategies, but I do acknowledge that more work needs to be done to meet this sort of objection.