



4-2005

Normative Failure in Blackburn's Ruling Passions

William R. Porter '05
Illinois Wesleyan University

Follow this and additional works at: https://digitalcommons.iwu.edu/phil_honproj



Part of the [Philosophy Commons](#)

Recommended Citation

Porter '05, William R., "Normative Failure in Blackburn's Ruling Passions" (2005). *Honors Projects*. 3.

https://digitalcommons.iwu.edu/phil_honproj/3

This Article is protected by copyright and/or related rights. It has been brought to you by Digital Commons @ IWU with permission from the rights-holder(s). You are free to use this material in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This material has been accepted for inclusion by faculty at Illinois Wesleyan University. For more information, please contact digitalcommons@iwu.edu.

©Copyright is owned by the author of this document.

Normative Failure in Blackburn's *Ruling Passions*
William R. Porter

—
421 Phoenix Avenue, Apt. 3
Bloomington, IL 61701
(309) 750-9773

—
Research Honors Project
Philosophy
Illinois Wesleyan University
April, 2005

INTRODUCTION

In *Ruling Passions*,¹ Simon Blackburn advances an ethical theory that welds his quasi-realism to a Humean-Smithean theory of moral sentiments. This paper concerns the latter Humean side of Blackburn's theory, specifically Blackburn's attempt to provide a normative ethical theory. This attempt largely involves getting over the tallest obstacle to any defender of Hume: the famous sensible knave problem.

I have three tasks. My first task is to examine Hume's original formulation of, and reply to, the problem posed by the sensible knave.² This task includes a treatment of a *prima facie* circularity problem in Hume's picture, which may be soluble with regard to the original sensible knave, but insoluble with regard to a variation or descendant of the sensible knave.

My second task is to examine the knave-related contemporary controversy between Blackburn and Christine Korsgaard. Blackburn's defense of Hume's reply to the knave consists primarily in his repudiation of criticisms leveled against Humean theories by Korsgaard in *The Sources of Normativity*.³ I will defend Korsgaard's argument that Humean theories cannot be normative because they fail to fulfill her requirement of transparency.⁴

Since I leave open the matter of whether Korsgaard's transparency requirement is itself a legitimate requirement on an ethical theory—I offer no arguments in support of it—my third and final task is to show that Blackburn's theory fails to be normative even when the transparency requirement is suspended or rejected. I will be assisted again in this last discussion by Korsgaard.

HUME: INTRODUCING THE KNAVE

In the discussion that precedes the appearance of the knave in the 2nd *Enquiry* Hume takes up the issue of our “interested obligation” to virtue (*E* 282). As Korsgaard says, Hume must overcome the worry “that morality might be bad or unhealthy for us” (*SN* 61). Morality should not be able to tell us what to do, unless it can tell us to do what is good for us. So Hume tries to show that from the point of view of self-interest, it is good for you to be a person who acts from motives approved by the moral sense.⁵

The problem the knave poses concerns the virtue of justice. As Postema⁶ points out, this is because the knave suspects that the demands of morality and self-interest are not congruent in the case of the artificial virtue of justice. No one can entertain knavish doubts about the natural virtues, on the other hand, because possessing those virtues (wit, for example) is its own reward, and thus exercising the natural virtues promotes self-interest.⁷ Justice, however, is another matter. On Hume's view, it is the *system* of justice that tends to promote the interests of society and individuals; and as societies expand, the benefits we receive from acting justly become less immediate. Moreover, justice occasionally requires a sacrifice of private or public interest.⁸ Such sacrifices cannot be motivated by a regard to immediate self-interest or benevolence, because they do not serve immediate self-interest or benevolence; and so in these cases the only normative standpoint available for motivating acts of justice is the standpoint of morality itself.

The difficulty should by now be clear: Hume must show that it is in our interest to be persons motivated by a sense of justice, even when being such persons may compel sacrifices of our interests. The knave throws the question directly back in Hume's face:

You say that we would endorse being a good person from the point of view of self-interest—well, what about in my case? Features of the knave's case require a special answer for this question.

Hume describes the character of the sensible knave in the following passage:

...a sensible knave, in particular incidents, may think, that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any considerable breach in the social union and confederacy (*E* 282).

The knave is a free-rider: he keeps the laws of justice when lawfulness serves his reputation, and breaks them in such "particular incidents" as do not return ill consequences—when a violation harms neither his self-interest nor the "social union."

This is the kind of problem that is virtually insurmountable for a voluntarist-contractarian theory such as Hobbes's, in which the moral authority of the sovereign rests in his power to punish violators of the law: when violations such as the knave's go undiscovered and unpunished, it seems the sovereign has no authority. Thus we cannot say to the knave that he should perform the just action since otherwise the king will have him beheaded (cf. *SN* 29-30). Saying this is not only theoretically tenuous—in the knave's case, it is not even *persuasive*, as the knave's example clearly stipulates that the king, or whoever, will know nothing of the offense.

The knave threatens the authority of a Humean theory as well, although for a different reason. If, as Hume contends, the good of justice derives from its general tendency to promote good (by securing property), then it seems the knave has no reason not to flout the demands of justice, given the proper opportunity; it may be that we surrender to the knave as soon as we admit even that there are exceptions to the general rule. By a perfectly accomplished piece of knavery, the knave hopes for the best of both

worlds, promoting his self-interest by the injustice, but without undermining the system of justice itself, nor the considerable interest he has in the continuance of that system.

Hume allows that there may be no correcting the knave:

If his heart rebel not against such pernicious maxims, if he feel no reluctance to the thoughts of villainy or baseness, he has indeed lost a considerable motive to virtue...(*E* 283).

However, Hume has available several responses. Postema offers a helpful survey of them. Drawing from *A Treatise of Human Nature* as well as from the 2nd *Enquiry*, Postema lays out “three related lines of argument in reply to the sensible knave.” Here are the first two,⁹ which, for Postema, are most easily dismissed:

(1) First, Hume maintains that because actions in such a practice [justice] are intricately linked, and success depends not simply on one's own action but on the coordinated actions of all the other members of the community, it is reasonable to believe that every violation of the rules genuinely threatens the survival of the system and thereby of society.

(2) Such exceptions [to the rules of justice as the exception proposed by the knave, Hume] insists, are permissible only if they can be formulated into a general rule and incorporated into the system of justice (*HR* 30-31).

Item (1) finds Hume denying that there are exceptions and holding that every act of injustice threatens the system of justice. Here, though, “Hume has overstated his case” (*HR* 31), Postema says. It appears simply false to claim that society cannot withstand a single act of injustice; society does withstand, quite often, the most severe acts of injustice, and “it has a great capacity to rebound from violations of even its most fundamental and widely shared principles” (*Ibid.*).

Item (2) fails to address the knave properly. The knave does not propose to replace the rules of justice with rules of his own; rather, “his argument is consistent with the conviction that, as a set of general rules for general compliance, the existing rules are the best conceivable” (*Ibid.*). The knave has no quarrel with the rules; furthermore, he knows (because he is a “sensible” knave) that if he wants to go on being a knave, he had

better obey these rules most of the time. His complaint is not to propose new rules of justice; it proposes an exception to rules he is generally willing to observe. (Remember this point. It will resound later against Blackburn.)

The third of Hume's replies is more sophisticated, and it is related to the first. Postema lists three conditions that would have to be met in order for the claim in (1)—that violations of the rules of justice threaten the whole system of justice—to be true:

(3.1) The actions of parties to the convention [of justice] are tightly locked in patterns of interdependence such that what each does depends greatly on what everyone else can be expected to do.

(3.2) The conventions are quasi-unstable, that is, while no party may prefer to deviate, *given* the conformity of all the others to the rules, nevertheless, deviation is preferred in the *absence* of sufficient *assurance* of general conformity.

(3.3) Violations of the rules of the convention are publicly detectable (*HR 33*).

The result of these three conditions is that the knave will have to keep his knavery secret, even while carefully maintaining the public appearance of being just. More perspicuously: the knave will endeavor to overturn condition (3.3), because it is the only one of these three conditions that lies directly under his control (cf. *HR 33*).

Postema says that Hume may argue that this strategy of meticulously guarded secrecy will prove disastrous for the knave. The knave will end up alienated at the core; in concealing a central principle of his life from his community, he cuts himself off from the community and so cuts himself off from *himself*. Since, in Hume's psychology, we identify ourselves by recognizing how others see us, the knave will have destroyed himself: he cannot see himself in others, because others do not see what he is.¹⁰

Postema chases Hume on this point through several twists and turns, and I do not have space for them here, so I will have to close my discussion of Postema's survey by observing this: after saying that he has reduced Hume's defenses to one last reply,

Postema suggests that that last reply is implausible. He has beaten back Hume on several fronts, but ultimately arrives at only a well-fortified skepticism about Hume's reply to the knave. I think Postema's arguments do yet allow that if Hume's claim (implausible, but merely implausible) that the knave "loses his character with himself" holds, then Hume's reply to the knave succeeds.

KNAVE ONE, KNAVE TWO, AND THE CIRCULARITY PROBLEM

This section continues my first task, examining some complications in our picture. I am preparing to show how Christine Korsgaard raises more serious problems for Hume, for she argues that Hume's theory fails even if it *can* show that the knave loses his character—and even if the knave gives in to duty and performs the just action in the exceptional case, when acting unjustly would benefit him without harming the system of justice.

I will present Korsgaard's arguments in the next section. In this section I will say a little more to show why they are needed. An apparent circularity problem can be solved with regard to the original knave; with regard to a variation of the knave, the circularity problem may be insoluble. As we will see, Korsgaard's arguments are remarkable because they threaten Hume's position regardless of the outcome of the circularity problem.

A kind of circularity "appears at first sight" (*SN 59*) to threaten Hume here. You may suspect that Hume's reply to the knave is circular if you take Hume to be saying that we should not be unjust because we disapprove of injustice. If this were really what Hume is saying, he would be begging the question, because such a response evades the

compelling possibility that the knave is simply not concerned with being just (although he *may* be so concerned, as Postema shows us). “Of *course* integrity will be cherished by honest people who feel the importance of it. But the sensible knave is questioning exactly that importance” (*Ibid.*).

Korsgaard explicitly allows that Hume's mechanism of sympathy can ensure that the knave act virtuously, despite his inclinations to the contrary and despite even his knowing that his transgression would go unnoticed. Sympathy imagines and *imports*, so to speak, the sentiments of others to bear on our deliberations; likewise sympathy enables the knave to imagine what the sentiments of others *would be if they knew of the knave's transgression*.¹¹ Thus, as Korsgaard says,

...the fact that *other people* will disapprove and dislike the sensible knave will be sufficient to provide him with feelings of disapproval and dislike of himself. Of course a knave will try to keep his knavish actions secret. But unless he is very hardened indeed, even the knowledge that others *would* hate him if they knew what he is up to will be enough to produce humility and self-hatred when he acts unjustly...So Hume's reply to the sensible knave is not circular (*Ibid.*).

Postema also notes several points in Hume's discussion at which Hume appears to beg the question. Recall that Postema's three conditions on Hume's first line of argument against the knave (which I've labeled {3.1-3.3}) give the result that the knave will have to keep his actions secret. A knave especially bent on knavery might well know that, and remind us that “knavery is not for everyone” (*HR* 34); the *appearance* of being just, the knave might say, accomplishes everything that the *reality* of being just would accomplish. So—and this is consistent with Korsgaard—if our knave is extraordinarily recalcitrant, he may not be motivated by the disapproving sentiments that infect him, preferring, after all, to keep his secrets. It is because the knave might press

the dispute this far that Hume is forced to argue that such secrecy *itself* is destructive of the knave's character (cf. *HR* 34-35).

Hume's reply to the knave would be circular if, as we might think "at first sight," there is no way of getting into the knave's head a sentiment that disapproves of injustice; if the knave is once and for all a person who cannot "feel the importance" of justice, it will beg the question to respond to him as though he does feel that importance. However, Hume's mechanism of sympathy can ensure that the knave will be visited by such a sentiment: the knave will internalize the sentiments of others who do disapprove of injustice, and so feel humility on that account. Moreover, as Korsgaard reminds us, Hume requires that we judge in sympathy with an agent's "narrow circle" (those people most immediately and frequently affected by the agent's conduct) and according to general rules. This requirement "brings a kind of objectivity" to the sentiments of others; there is a "convergence of sentiments" (*SN* 55). When the knave internalizes the gaze of others, he internalizes a sentiment that disapproves of injustice univocally.

For Blackburn, the solution to the problem ends here: if the knave's sympathetic apparatus is operating properly—and it should, insofar as he is a "sensible" knave—then he will be provided, under sufficient societal pressures, with motivations for doing the right thing, however grudgingly. If it turns out that this kind of reluctant virtue marks the regular pattern of his behavior, then we have the best that may be hoped for in his case, and we may be inclined to redouble our efforts to educate our fellows toward a genuine love of virtue. Perhaps we might even use the knave as an example to the young: *wouldn't it be easier for him, we could say, if he enjoyed being so law-abiding?*

Blackburn distinguishes two kinds of sensible knave, and is far more alarmed by the threat posed by the second. Henceforth I will refer to the original knave as “Knave One,” and to Blackburn’s second knave as “Knave Two.” Blackburn notes the example of Rousseau’s Spartans, who kept a strong society among themselves but dealt with their neighbors by violence, treachery, and deceit. He calls this “Foreign-Office variety” of the knave—Knave Two—the “less tractable” and “unfortunately also more common” problem (*RP* 211).

Postema shows us that Knave Two could be the evil descendant of Knave One:

The present argument [Hume’s] does not show that it is in the knave’s interest to follow the dictates of justice in those cases in which injustice will be undetected and the victims are sufficiently psychologically (or socially) distant. This is a serious problem for Hume. For the special attraction of justice is its promise to *extend* the patterns of cooperation typical of the close circle of the family to larger and less personal social contexts in which natural relations of trust and close contact are weakened. Now it appears that justice can be shown to be rational only if the circle of persons to whom justice is owed does not extend very far beyond the circle of family and friends (*HR* 36).

Humility may stop Knave One short of acting unjustly when the injustice harms people with whom he strongly sympathizes. If he then turns his opportunistic injustice against distant victims with whom he does not strongly sympathize, he transforms into Knave Two.¹²

On Hume’s theory, we enter a system of justice by modeling, in ever-wider scope, the kind of familial cooperation into which we are born. This system of cooperation will serve our interests much in the way that family members help each other in primitive or tribal settings—and I don’t intend to speak pejoratively here, since such settings can be hypothetical (cf. *HR* 25-26). We extend the system of justice, because it promises to promote our interests—and we do this by extending our sympathy, creating a broader and broader community.¹³

So one hopes, at any rate. But now Knave Two threatens to send the whole process into retreat. We may supply Knave One with a motivation toward justice by infecting him with feelings of humility; but what if the proposed piece of knavery, Knave Two's, involves making victims of distant strangers, with whom neither the knave nor his community sympathizes strongly? History's darkest tales often involve men made famous by just this sort of knavery—take the Spanish *conquistadors*, for instance—who may have been virtually impervious to sentiments sympathetic to their victims. For knaves like these, sympathy may be too weak, with regard to their victims, to be active. Circularity may indeed be ineradicable in the case of Knave Two, because not even Korsgaard's counterfactual sentiments can come to the rescue as they did in the case of Knave One: a *conquistador* in a community of other *conquistadors*, proposing to murder and rob an Aztec emperor, could not feel humility by considering what others *would* think of the action—most likely, the others would approve it.

If one thinks, as Blackburn seems to think, that the problem begins and ends with the technical challenge of sneaking a sympathetic sentiment into the knave's head, then Knave Two will indeed pose the less tractable—and possibly the intractable—circularity problem. This line of thinking also explains why the final chapter of *Ruling Passions*, in which Blackburn purports to address the issue of authority, is primarily an extended repudiation of ethical relativism. Knave Two raises the spectre of ethical relativism, because his example suggests that, should two antagonistic societies full of Knaves Two come into conflict (as in, say, the horrific Rwandan genocide), there will be no normative ground to stand on in resolving the dispute.

We turn now to Korsgaard, because she, alone among these philosophers, argues that normative problems still haunt the case of the sensible knave—of whichever variety—even if he can be instilled by sympathy with the proper sentiments. The problem of moral authority is a much more intimate problem than the macrosocial problems of ethical relativism; and the knave of the domestic variety poses a problem less tractable than Blackburn supposes.

KORSGAARD'S TRANSPARENCY REQUIREMENT

I move now to my second task: a discussion of the contemporary knave-related controversy involving Korsgaard and Blackburn.

The following arguments from Korsgaard apply *whether or not the knave is infected by sympathetic sentiments*. If the knave never even acquires such sentiments (as Knave Two may not), then Hume's project dies before Korsgaard can kill it. If the knave does acquire such sentiments (as we suppose Knave One will, at least), Korsgaard can show that moral authority still remains at risk. This is why these arguments are devastating for Blackburn's treatment of the sensible knave.

According to Korsgaard, a normative ethical theory must meet a requirement of *transparency*: that is to say, the application of a moral theory must survive the agent's knowledge of that theory. If knowing the truth about morality would destroy the agent's motivation to be moral, such a morality cannot be normative, because the agent cannot in this case act morally without believing that her morality is *unjustified*. Blackburn's failure to meet the challenges presented by Korsgaard is due largely to the fact that his

responses seem to ignore the transparency requirement, and without arguing for its illegitimacy.¹⁴

To make the pertinence of the transparency requirement to our considerations especially clear, Korsgaard asks us to imagine “a slightly more attractive version of Hume’s sensible knave”:

Our knave is the lawyer for a rich client who has recently died, leaving his money to medical research. In going through the client’s papers the lawyer discovers a will of more recent date, made without the lawyer’s help but in due form, leaving the money instead to the client’s worthless nephew, who will spend it all on beer and comic books (SN 86).

What adds to the fun is that our lawyer “is also a student of Hume, and believes the theory of the virtues that we find in *A Treatise of Human Nature*” (SN 88). This feature makes our case a dramatic test of the transparency requirement: our agent is fully aware of the nature of the morality that guides her deliberation, and we must see whether she will feel justified if she does the right thing.

It turns out that our lawyer will not have such a feeling of justification. For, as a student of Hume, she knows not only that she would disapprove of herself if she executed the invalid will, but also *why* she would disapprove of herself: because the kind of deceit involved in executing invalid wills has the general tendency to harm the social order. In this case, however, her deceit would have no harmful effects; in fact, we should say that it would have very helpful effects. As Korsgaard says,

The lawyer believes that her disapproval of this action [executing the invalid will] depends on the fact that actions of this kind usually have bad effects which this one does not have. It is almost inconceivable that believing this will have no effect on her disapproval itself (SN 87).

If Hume is right, the lawyer may find that she cannot destroy a valid will without intense feelings of humility or self-hatred. These may or may not be strong enough to cause her to desist. But even if they are there will have been normative failure. The lawyer does not believe that the claims her moral feelings make on her in this case are well-grounded. If she could cure herself of them then that is what she would do (SN 88).

Blackburn offers a response to Korsgaard's Humean lawyer, but it is inadequate. He begins with a few contempting remarks characterizing the scenario as a "magical" case, one from "philosophical fairyland" (*RP* 233 and 235, respectively): any real lawyer, in the real world, would face innumerable complexities that this scenario must, as a mere sketch, exclude.

These remarks should be ignored; they simply dodge the point. Hume himself posited the problem of the sensible knave, and so we should expect a defender of Hume to do better than to hand-wave away the very epistemological condition that makes the problem so interesting and difficult. The epistemological condition is that the knave *knows* that a particular unjust action will bring consequences utilitarianly preferable to those promised by the just action that is its alternative—and of all the scenarios answering to this condition, Korsgaard's seems most charitable for its being rather notably unmagical.¹⁵

Ignoring, then, Blackburn's first criticism of the Humean lawyer, we should look at his more mannered and thoughtful response, which attempts to refute Korsgaard's contention that the lawyer would rid herself of her "moral feelings" if she could. I have placed "moral feelings" in quotation marks because of the sly inaccuracy of the following (notice where I have added emphasis to Blackburn's text):

So the key question is why the Humean lawyer should wish she could cure herself of her *disposition* to hate injustice—the disposition that is causing her to hate it even in this case, where the injustice will do good. We might have found this question difficult had we not been through the discussion of the previous chapters....Among the motivational states Hume would approve in a lawyer will be the desire to act as her role demands; respect for the wishes of the deceased; and the pride or at least peace of mind that comes of acting only in ways that stand up to public scrutiny (*RP* 234).

In the next section, I will examine the "discussion of the previous chapters" to which Blackburn alludes in this passage; we will see that it has problems of its own.

In the meantime, we should notice that what Blackburn says here *is* mostly correct, as far as it goes: the Humean lawyer, properly speaking, will not wish to cure herself of her *disposition* to hate injustice, because, as a good student of Hume, she knows that this disposition generally serves public and private interest. But this is not, as Blackburn says, the “key question.”

That’s because saying “disposition” where Korsgaard quite meaningfully does not—she says “moral feelings”—is a crucial mistake. Blackburn is right: the Humean lawyer will not wish to cure herself of her disposition to hate injustice. What she will wish to be rid of, rather, is the *operation* of this disposition *in this particular case*. This case is exceptional: features that uphold the general rule describing the operation of the disposition—the features that in most cases explain the tendency of the disposition to promote the societal good—are not present. And these are the very features that allow the lawyer to feel, *in most cases, but not in this one*, that the claims made on her by morality, or by her role, are justified.

Recall Postema’s second Humean reply to the knave: the argument that exceptions are impermissible unless they can be assimilated to a general system of rules. Postema points out in that argument a mistake similar to the one I point out here: just as the knave does not challenge the existing rules, and indeed allows that they may be the “best conceivable,” the Humean lawyer does not wish to change her disposition to hate injustice. This is another reply that fails to address the knave properly.

Korsgaard in fact anticipates Blackburn’s answer, because Hume made the same answer. In her discussion of the lawyer-knave, Korsgaard tells us what is wrong with it:

Hume has a defence against this point, but it is a defence of the wrong kind. Consider once more the original sensible knave. What does he lose by his knavery? According to Hume, he loses his character with himself, his pleasing sense of self-worth...[and] this

does not depend on his moral beliefs, or on whether he endorses the claims of morality. Since sympathy makes him see himself through the eyes of others, this will happen anyway. But this is exactly the problem (*SN* 87).

The “problem” is that although the knave’s knavish actions may cause him to feel humility, he will not feel that this humility is justified, because nothing in the case recommends a virtuous action for the regular reason or according to the general rule. Moreover, even if the knave could be moved by sympathy, most or some of the time, to perform the virtuous action, he (or she, in the case of the lawyer) would still lack a feeling of justification. Korsgaard here completes the criticism that Postema was unwilling to pursue further: even if, as suggested by Hume’s third line of reply, the knave will lose his character with himself, he will feel that this “loss of his soul” is ill-grounded. This lack of justification is what Korsgaard calls “normative failure,” and it poses a threat that Blackburn’s theory cannot surmount so long as the transparency requirement stands.

NORMATIVITY, PRIVACY, AND BLACKBURN’S *A PRIORI PRINCIPLE OF INTERPRETATION*

We turn now my third and final task, and to “the discussion of the previous chapters” to which Blackburn alludes above. What goes on in this discussion? How, exactly, does this discussion assist Blackburn’s position—and does it?

The discussion Blackburn has in mind concerns a case of his own, which he calls “structurally the same” (*RP* 235) as the case of Korsgaard’s lawyer-knave. The case is that of a soccer referee who faces an opportunity to make a deceitful call that might bring certain desirable consequences—it might, say, benefit the home team and delight the home crowd of a hundred thousand, a rowdy and devoted bunch of hooligans who have been known to riot after a loss. The alleged structural similarity between the deliberative problems faced by the Humean lawyer and the soccer referee seems to be this: each is

weighing, against an ostensibly duty-bound alternative, an act of deceit that is utilitarianly preferable. The question for Blackburn, as a defender of Hume, is how to make room in his theory for the recommendation of the dutiful action, despite Hume's seemingly utilitarian leanings.

Interestingly, Blackburn's first impulse in the treatment of the referee case is, as in the case of the lawyer, to sneeze at the epistemological conditions that define the scenario. Just as Blackburn contends that the lawyer would not be so likely to know that the execution of the invalid will would bring happy consequences, he contends also that the referee would not be so likely to know just which false call would be the one utilitarianly worth making. The thrust of this point, then, seems to be that the referee ends up served best, even in utilitarian terms, by somehow forgetting the possibility that a false call might please the crowd. Consider this passage from Blackburn, in which we might consider the Humean lawyer or the soccer referee as an example of a "participant," and professional law or refereeing, respectively, as an example of an "institution":

For what is to stop the participant from being alert to situations in which the general good is indeed furthered by his breaking a rule? Surely the very story we tell about the good generated by the institution must allow that there should be such cases, and that the participant would do better to exploit them. For, *ex hypothesi*, they are cases in which the good is furthered by the infringement. So we can have the general good plus the surplus good created by occasions of departure from the rules... The mistake in this tempting line of argument is in supposing that we *can* have the general good plus the surplus. The incoherence of this as a general recommendation is already apparent. For if it were generally known (for example) that referees were disposed to give false decisions when they judged that in that way the interests of the spectators and players would be better served, then, as we have seen, the entire structure collapses (*RP* 42).

Here Blackburn is arguing that a consequentialist doctrine can to an extent determine the thoughts of genuine participants in a consequentialist institution. It can forbid certain considerations from the participant's deliberation; it can even—so Blackburn hopes—forbid the participant from deliberating as a consequentialist. But this approach

falls easy prey to the transparency requirement: it requires that our referee and our lawyer be unaware of the ends they would serve by performing their offices; a consequentialist institution that requires its participants not to deliberate as consequentialists is not a transparent institution.

The result is that if, as Blackburn says, the cases of the Humean lawyer and the soccer referee really are structurally the same, then the latter falls prey to the same arguments from the transparency requirement that dissolve normativity in the former. If they share the same structure, they share the same defect. Referees and lawyers alike represent failures of Humean ethics so long as the transparency requirement stands.

By the way: I think that the two cases are, in at least one important respect, *not* structurally similar, for the Humean lawyer considers a transgression that would be kept secret, while the referee considers a transgression that would take place before an audience of thousands. So much the worse for Blackburn: if the analogy is a good one, his arguments regarding the referee fail so long as we recognize the transparency requirement; and if the analogy is not a good one, then referee-related arguments can't be put into service in reply to the Humean lawyer. Blackburn's last hope, then, is that we ignore the transparency requirement.

So in charity to Blackburn, let us suppose, however dubiously, that Korsgaard's transparency requirement is not a legitimate requirement on an ethical theory. Even so, I think Korsgaard would find Blackburn's theory unsatisfying in another way as well. To open this discussion, we should take broaden our discussion of ethical "roles" beyond just professions or offices, such as law and refereeing; we may now include regular behavioral patterns, or what Blackburn often calls "contingent profiles of concern."

According to Blackburn, these profiles come parceled with a normative dimension: an agent's pattern of behavior will allow us to "read back" from her actions her concerns, and those concerns in turn can tell us what that agent *should* do.

Now, the concept of "concern"¹⁶ is rather narrow. For purposes of illustration, we can start more broadly, with desires. Let's suppose that you desire to eat an apple. I may learn about your desire to eat an apple by your engaging in any number of characteristic behaviors—there is an apple nearby, say, that you've been eyeing intently for some time, or you've just said, "I really could go for an apple." Now, I can expect that you will eat an apple, and will be surprised if you do not eat an apple—this much is simple.

Blackburn's view, though, makes the further step of describing my expectation as imparting (or at least *belying*) a normative dimension to our situation: my expecting that you will eat an apple amounts to my believing that you *ought* to eat one, or that your eating an apple would "make sense." There is thus, on this analysis,

an assimilation of the normative and the causal order. We know what a desire is by knowing what it would make sense to do in the light of having that desire; but then we know whether someone has the desire by seeing if this light is one that makes good sense of what they [sic] do (*RP* 58).

And all this holds for the narrower concept of concerns, as well. Over time, the prudent actions of a prudent person inform us that she is *concerned* to be prudent, and so we may come to expect that she perform—and think, therefore, that she *should* perform—more prudent actions.

Informing this model is what Blackburn calls the "a priori principle of interpretation," or API. Here is Blackburn's statement of this principle:

It is analytic that creatures with beliefs, desires, and other states of mind, behave in ways that (best) make sense (and not in ways that make no sense), given those states of mind (*RP* 55).

It should be clear how such a principle delivers the above analysis of desires and concerns. Can the principle, though, have further implications for our discussion of the knave? We should consider API's bearing on the problem of the sensible knave, since API figures so prominently in the "discussion of the previous chapters" of *Ruling Passions* that furnishes—says Blackburn—an obvious and total repudiation of Korsgaard.

One might suspect that API gives us a back door to moral authority by imposing what we could call "normatively narrow" psychologies on moral agents: psychologies that, specific to an agent's role or office, exclude certain considerations from the agent's moral deliberations *so long as that agent may be accurately described as inhabiting that role or office*. So, concerning our lawyer and our referee, we might be able to say something like this: so long as lawyers and referees are performing *as* lawyers and referees—and that is to say, as people concerned with being good lawyers and referees—they will perform as their offices require, and feel justified in doing so. Otherwise, they are just not performing as lawyers and referees; it is analytic that to perform as a referee is to perform according to certain rules. There are only so many mistakes one can make as a referee without forfeiting the name. As Blackburn says,

Consider the example of game-playing. Here, too, there is a definitive normative order: a game is defined by its rules. There is a limit to the extent to which people can fail to conform to the rules. A rule may be broken now and again, but systematic and acknowledged breaking of the rules becomes not that, but a change of the game. Yet it is largely an empirical matter which game people are playing. Their behaviour tells us which patterns they *do* conform to and that in turn tells us which game they are playing: that is, which are the rules to which they *ought* to conform (*RP* 57-58).

But on this approach, what regular behavioral profile could fail to count as a role, station, or office? So far as I can tell, Blackburn nowhere in *Ruling Passions* addresses this problem: that API confuses our *expectations* of agents (relative to their roles) with the normative *demands* we would make of them.

These are really not the same, but they cannot be distinguished under API, because although API may explain the way we expect other agents to act in accordance to their roles, it cannot render any judgments or offer any recommendations as to such roles *themselves*. You may, in certain cases, decide which action is best by considering your role, but you cannot decide which role is best by appealing to your role. Consider this example:

Mr. X is a swinger.¹⁷ He considers himself a ladies' man; the feeling of manliness he derives from swift and short sexual conquests is genuinely constitutive of X's sense of identity: if a trusted friend asks him to define himself, he defines himself primarily as a user of women. Now, X's acquaintances may, given X's deplorable track record, expect X to spend next Saturday night, as so many before, at some nightclub, trying to beguile some luckless woman into sleeping with him.

But—unless, as I suppose is plausible, these acquaintances share and endorse X's Bacchanalian tastes—we shouldn't think that these expectations are necessarily to be identified with what X's acquaintances think X *ought* to do. The word *ought* has no place here: we don't speak of philanderers as if they *ought* to philander, just because doing so is their habit. We say rather that they *ought* to throw over that role entirely.¹⁸

Korsgaard, I think, would reject Blackburn's claim that there is an "assimilation of the normative and the causal order." Our inability to distinguish the two under Blackburn's account is precisely what disturbs us in the case of Mr. X: we don't want to give in to the thought that X's relationship to his role as a

philanderer is a causal relationship; we want to say, rather, that he can *disobey* the “demands” of his role and reform himself. If we are especially evangelistic, then for the sake of X’s immortal soul, we insist on that.

In the fourth lecture of *The Sources of Normativity*, Korsgaard demonstrates the importance of a distinction between the normative and the causal through a treatment of Wittgenstein’s private language argument. Wittgenstein proved that there cannot be a “private” language, a language which cannot logically be understood, except by its only speaker. Wittgenstein’s argument proceeds by *reductio ad absurdum*; he begins by imagining what an allegedly private language would be like, and presents the following scenario.

Your private language consists of your naming a sensation whenever it occurs. You can’t call it by the name of any familiar sensation, because in that case you would be able to communicate it to us, and your language wouldn’t be private. Instead, you call the sensation “S.”

Wittgenstein argued that there couldn’t be any such language. One way to understand his argument goes like this: meaning is relational because it is a *normative* notion: to say that X means Y is to say that one ought to take X for Y; and this requires two, a legislator to lay it down and a citizen to obey. And the relation between these two is not merely causal because the citizen can disobey: there must be a possibility of misunderstanding or mistake (SN 137).

There *must* be a possibility of misunderstanding or mistake, because, as Wittgenstein has it, there can be no possibility of correctness without the possibility of incorrectness. Consider what happens when you name your sensation “S” *for the first time*. As Wittgenstein writes,

this process [of naming the sensation ‘S’] brings it about that I remember the connection *right* in the future. But in the present case I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we cannot talk about ‘right.’¹⁹

Since that first decision could not have been incorrect—you could have given the name “S” to any sensation—it could not have been correct, either. Correctness requires the possibility of incorrectness; language requires both possibilities, and thus there can be no private language.

According to Korsgaard, “we can make a parallel argument against private reasons: reasons are relational because reason is a normative notion: to say that R is a reason for A is to say that one should do A because of R: and this requires two, a legislator to lay it down and a citizen to obey” (*SN* 138).²⁰ Blackburn’s case, as we have it, imputes to moral agents just the sort of “private reasons” that are under attack here. What an agent ought to do is to be settled by an appeal to the nature of that agent’s role or profile of concern. But the profile of concern bears a causal relationship to the action, and so what ought to be done is isomorphic with what is expected to be done.

The problem with this approach is much like the problem you face when you privately name a sensation “S”: there is no possibility of a mistake. And yet there must be a possibility of a mistake: surely we can abandon our roles if we wish; and surely we *should* abandon our roles when it is right to do so. Since API will describe your behavior as “making sense” (given your profile of concern) *no matter what you do*, it “cannot talk about ‘right.’” API interprets every action of every agent as “making sense”; thus obedience and disobedience to your role “make sense” alike—and thus the interpretation is otiose.

CONCLUSION

I have presented two lines of argument along which Korsgaard can successfully defend from Blackburn's criticism her general thesis that Humean moral theories fail to provide a source of normativity.²¹

If the transparency requirement stands as a legitimate requirement on an ethical theory, then a Humean theory fails in the case of the knave. Even if the sensible knave can be motivated by counterfactual sentiments—thoughts of what others would think of him if he were to do the unjust action—he would still suspect that his duty-motivated action was unjustified. Such an example illustrates what Korsgaard calls “normative failure”: to fulfill the transparency requirement, an agent must be able to act in the light of full knowledge of morality and feel justified in doing so.

If the transparency requirement is suspended or rejected, Blackburn's theory still fails to provide a source of normativity. It cannot provide a source of normativity because it is founded in a principle, API, that presumes the isomorphism of the normative and the causal. Korsgaard's deployment of Wittgenstein's private language argument shows that normativity cannot survive a complete alignment with causality. API fails, then, to provide any normativity because it cannot describe role-playing agents as making any mistakes.

I think these two analyses are most illuminating when considered as an interdependent pair; in a sense, Blackburn's theory fails because it seems not to recognize the relatedness of these points. In the end, it is perverse that Blackburn asks us to remember API in support of his response to the soccer referee, because API holds that “there is an assimilation of the normative and the causal order.”

But it is precisely because this is not the case that the referee and the Humean lawyer and the sensible knave pose compelling normative problems. If there were truly an assimilation of the normative and the causal order, then a referee would stand in a causal relationship to his role; the role would *cause* him to do things, and we would not have to wonder about whether he should choose to break a rule. API is not a governing principle; if it were, there would be no role-inhabiting sensible knaves at all. The relationship between an agent and her reasons for action, role-related or otherwise, can be normative only insofar as it is *not* merely causal.

¹ Oxford: Clarendon Press, 1998. Hereafter cited in text as *RP*, followed by corresponding page numbers.

² I draw mainly from Hume's 2nd *Enquiry*, edited by L.A. Selby-Bigge and revised (1974) by P.H. Nidditch. 3rd edition. Oxford: Oxford University Press, 1999. Hereafter cited in text as *E*, followed by page numbers.

³ Cambridge: Cambridge University Press, 1996. Hereafter cited in text as *SN*, followed by page numbers. Responses to Korsgaard's lectures by G.A. Cohen, Raymond Geuss, Thomas Nagel, and Bernard Williams also appear in this volume, but I have made no use of these here.

⁴ A term Korsgaard uses technically; a more thorough explication of the notion of transparency follows.

⁵ As many philosophers have pointed out, Hume is not on this count an ethical or psychological egoist. Hume does not follow Hobbes and argue that we should exercise the virtues from self-interest because it is the only motive we can have; nor does he argue that we should exercise the virtues just to feel good about ourselves. Hume argues, rather, that from the point of view of self-interest it is good to be a person who is morally good. There is a difference between saying with the egoists that virtue is *directed* at your interest, and saying, with Hume, that virtuous interests are good interests to have. A morally good person acts from motives approved by the moral sense; she in turn approves of her virtue all over again when she regards it from the point of view of self-interest. As Korsgaard puts it, "it is clear that Hume is not saying that we should perform *particular* virtuous or obligatory actions because it serves our own interest to do so. He is saying that it is in our interest to be *people who practise virtue for its own sake*" (*SN* 60).

⁶ In "Hume's Reply to the Sensible Knave." *History of Philosophy Quarterly*, Vol. 5, No. 1, 1988. pp. 23-40. Hereafter cited in the text as *HR*, followed by page numbers.

⁷ Hume writes: "That the virtues which are immediately *useful* or *agreeable* to the person possessed of them, are desirable in a view to self-interest, it would surely be superfluous to prove. Moralists, indeed, may spare themselves all the pains which they often take in recommending these duties. To what purpose collect arguments to evince that temperance is advantageous, and the excesses of pleasure hurtful. When it appears that these excesses are only denominated such, because they are hurtful..." (*E* 280). Temperance is a natural virtue, and as other natural virtues it is *defined* ("denominated") quite specifically by its tendency to our interest; contrarily "excesses of pleasure" have their tendency against our interest built into their definition—if they weren't bad for us, they wouldn't be called "excesses." So, Hume says, a proof that the natural virtues are in our interest is indeed superfluous; the proof is already there, in our definitions or denominations of those virtues.

⁸ It is easiest, perhaps, to imagine cases—including the knave's, certainly—in which an act of justice runs counter to self-interest. Importantly, though, Hume notices that acts of justice may also run counter to the public interest. In the case of the "seditious bigot," a creditor repaid by "a man of merit, of a beneficent

disposition...[,] the public is a real sufferer." *A Treatise of Human Nature*, edited by L.A. Selby-Bigge and revised (1978) by P.H. Nidditch. 2nd edition. Oxford: Oxford University Press, 1990. p. 497.

⁹ Here, as well as in the subsequent presentation of items {3.1-3.3}, my text reproduces Postema's exactly; however, the ordinal numbers assigned to these items are my own.

¹⁰ See esp. *HR* 34-35.

¹¹ The imaginative operations of the knave's sympathy in providing counterfactual sentiments (what he *would* feel) stand in close analogy, I think, to the imaginative operations of the common point of view in Hume's treatment of another special case, that of "virtue in rags." We can love virtue in rags (that is, virtue that fails to deliver the usual consequences of virtue) because, from the common point of view, we see that such virtue *would* bring about the usual good consequences if it were exercised under the usual circumstances. The knave's imaginative sympathy supplies him with counterfactual sentiments that, if all goes well, provide an effective motive to the right action and thus precipitate good consequences; our regarding virtue in rags from the common point of view provides us with counterfactual consequences that, if all goes well, arouse the proper sentiments of approbation.

¹² The strength or weakness of sympathy owes to Hume's associative principles: sympathy, in particular, is enmeshed in considerations of *resemblance* and *contiguity*. Thus you sympathize more readily and strongly with people who are close to you and who are like you. And thus Cortez the Spaniard may have been nearly incapable of sympathizing with Aztecs in strange clothing, with strange skin, speaking a strange language, who had grown up on a strange continent halfway around the planet.

¹³ Again I must mention that Hume is not an egoist. Systems of justice have their origin in our self-interest, but as they mature, they gather a sort of moral momentum: eventually we may love justice for its own sake, and so further extend the system of justice for moral reasons instead of for self-interested reasons.

¹⁴ In setting out the transparency requirement, the example Korsgaard has in mind of an ethical theory that most plainly fails this requirement is Bernard Mandeville's, in which morality is an instrument of political power: our leaders sing the praises of virtue to enshrine their own power and wealth by keeping a stupid public humble and subservient. If you or I, members of the public, were to become suddenly less stupid, our motive to virtue would disintegrate. See *SN* 8-9.

¹⁵ I feel further inclined to remark Blackburn's bad manners on this count. His contempt for Korsgaard's "magic" seems especially misplaced in a chapter of *Ruling Passions* that asks us to consider the example of a fellow named "Citizen," who confronts his doppelgänger in a competitive economic negotiation.

¹⁶ Blackburn and Korsgaard more or less converge on their use of the term "concern," which we may define as a long-standing desire secured by a second-order desire: the agent desires not to lose her concern, as she considers the concern to be constitutive of her sense of herself. (Thus concerns, in this use, are distinguished from merely persistent desires which we may desire to be rid of, such as various harmful addictions.) For Korsgaard, concerns are therefore the constituents of one's practical identity: the agent suspects that, were she deprived of them, she would not really be herself.

¹⁷ A Platonist's objection might simply deny that the vicious man has any moral identity whatever, on the basis that moral identity requires a certain harmony of the soul; the vicious man or tyrant, on this view, is somehow "at war with himself." I feel strongly that such objections should be disallowed here. Experience shows—lamentably, I admit—that Mr. X's brand of vice at least is all too psychologically sustainable. I have deliberately avoided treating, say, serial killing as a role, as it would likely be objected that serial killers are somehow schizophrenic and identity-less. Mr. X is certainly not a schizophrenic.

¹⁸ Mr. X's case is especially illustrative because what we expect of him is *opposed* to what we think he ought to do. Notice, though, that something might continue to haunt us even when what we expect of an agent is *aligned* with what we think he should do: do we really want to say that a philanthropist should be charitable just because we *expect* that he will be charitable?

¹⁹ Wittgenstein. *Philosophical Investigations*. Translated by G.E.M. Anscombe (New York: Macmillan, 1953). Section 258, p. 92.

²⁰ On Korsgaard's account the two that make for a genuine source of normativity are the "thinking self" (the legislator) and the "acting self" (the citizen): thus, morality is a matter of self-legislation, of giving laws to oneself.

²¹ I should mention that Blackburn, for his part, offers criticisms of Korsgaard's positive account; this paper is only half of Korsgaard's defense. The other half is a topic for another paper.