

The fugitive thought

Richard Joyce

[This is the penultimate draft of the article that appeared in *Journal of Value Inquiry* 34 (2000) 463-478.]

Moral imperatives are claimed to be inescapable. The moral felon who convinces us that he desired to commit his crimes, that he had no desires that the actions thwarted, does not incline us to withdraw our judgment that he did what he ought not to have done. We do not permit him to evade his moral culpability by citing unusual desires or interests. This thesis of moral inescapability seems familiar and yet is notoriously difficult to make sense of. Philippa Foot calls it “the fugitive thought,” and argues that there is no coherent thought to be brought into the light, but thinks that we could carry on moral discourse purged of claims of inescapability.¹ Her pessimism about locating the fugitive is justified, but moral discourse may well not survive its elimination. Inescapability—our tendency to morally condemn the criminal regardless of his desires or interests—lies at the heart of our moral framework; indeed, we might well think that it is the whole point of having a moral language. For this reason our moral discourse is hopelessly flawed in the sense of there not being an acceptable explication of the central moral concepts, though it is not without practical merit. Participating in moral discourse may be warranted in pragmatic terms, despite its defects.

Inescapability finds its expression in familiar moral rhetoric such as: “You must not do that,” “I had to do it,” or “You really ought to do it, whether you feel like it or not.” It is natural that an examination of the plausibility of inescapability becomes a discussion of reasons. Imagine that an unregenerate criminal asks of a particular crime that he desired to commit “Why should I not do so?” Defenders of moral inescapability must be able to say something more than “You must not because it is forbidden.” To say no more than that is to concede the debate immediately to the skeptic; to try to say more is to attempt to provide the felon with a reason. John Tilley notes that there would be little point in being able to convince a person “that he is morally required to do *D*, if he can grant what we say and still ask for a reason to do *D*.”² It is not the status of the notorious question “Why should I be moral?” that is under examination; at least, in so far as that question is taken, as it usually is, in a general, *de dicto* manner. We are examining a hypothetical question concerning a particular moral action, such as an act of promise-keeping. A promise-breaking culprit may ask: “Why should I have kept that promise?” “Why should I be moral?” threatens to collapse into a triviality: “Why ought I do what I ought to do?” In that case, that it is a puzzling question comes as no surprise. In contrast, “Why should I have kept that promise?” is not trivial. The thought that we ought to keep promises is much more straightforward and central to our moral discourse than the peculiar thought that we ought to be moral. If it is claimed that we ought to keep promises, then the question “Why?” is utterly reasonable. If a good reason cannot be given, then we have no business asserting that the promise-breaker did something that he ought not to have done.

There are any number of good reasons that can usually be given for keeping promises. Promise-breaking is likely to incur mistrust, the wrath of the sovereign; promise-keeping gives us, in Hume’s words, “inward peace of mind, consciousness of integrity, a satisfactory review of our own conduct.”³ It need not be self-interest that grounds reasons for promise-keeping: perhaps I simply have a desire to see a fellow human being, to whom I have made a promise, avoid the harms of mistreatment. But all such reason-providing circumstances are contingent. We can easily imagine a promise-breaker who is specially situated, or has

peculiar desires. Perhaps he does not care about mistrust, punishment, or inward peace of mind; or perhaps he is so situated that he avoids or obtains these things in spite of his hurtful behavior. A familiar philosopher's fairy tale contains a Lydian shepherd who finds a ring of invisibility.⁴ If it helps, we will allow the moral felon to be similarly endowed. Hume would have us believe that the shepherd, despite avoiding the more tangible costs of law-breaking, still incurs the important loss of fellowship, but the argument is inherently weak. The criminal may participate in a sincere and caring manner in his local community, and would not dream of cheating his friends. It is only upon a neighbouring community that his harmful activities are visited.

If for all of us the Hobbesian and Humean reasons to behave ourselves are perfectly plausible, then does it matter whether Plato's far-fetched character would have a reason to refrain from lying, stealing, and even killing? It does. If we allow that the imaginary felon has no reason to refrain, then we can hardly seriously hold that he ought to refrain, and so the ought-statement would be escapable. But moral "ought"s are intended to have an inescapability; moral language is exactly that with which we would hope to censure the Lydian shepherd. There may be good reasons for me to keep promises, but we could not say that I morally ought to keep promises, if the force of that prescription is that I would have reason to do so even if I had such a ring. If we cannot provide the Lydian shepherd with a reason to keep promises then our actual moral discourse is in jeopardy. The worst case scenario is that there is no such thing, and never has been, as a true moral judgment.

We have taken it for granted that there is a connection between it being the case that a person, *S*, ought to do something, ϕ , and *S*'s having a reason to ϕ , and this needs examination. Let us look first at some non-moral "oughts." There is an "ought" of prediction, as in "Henry ought to be here any minute." Here it clearly does not follow that Henry has a reason to be here at any minute. This "ought" does not have reason-giving qualities because it can be applied to inanimate objects: we say that there ought to be an avalanche soon, due to the warming snow, but clearly the snow does not have any reasons.⁵ The "ought" that we are interested pertains to actions. Foot's agenda-setting paper introduced us to what might be called a weak categorical imperative, an imperative that is not withdrawn when it is discovered that the person addressed has no interest in the prescribed action, and no desires or needs that the action is likely to satisfy. This is the case for the rules of etiquette. When I discover that Mike has no interest in table manners, and cares nothing for my feelings, I do not retract my imperative of etiquette: that he ought not to spray me with bread while he eats. Suppose Mike has an excellent reason for having appalling table manners. Perhaps he stands to win a large bet if he annoys me over lunch. Then we might well say not only that he has a reason to speak with his mouth full, but that he has no reason not to. Foot's point is that despite his possibly having no reason to respect etiquette, an imperative of etiquette, "Mike ought not speak with his mouth full," applies. This "ought" might be called an institutional "ought."

Foot does not think that this weak categorical imperative provides the basis for inescapability. Inescapability she identifies with the putative reason-giving quality of morality. Moral imperatives do not continue to *apply to us* regardless of our motives or interests in the way that the rules of etiquette or chess do. We have reason to *follow them* regardless of our motives or interests. We might put the point in terms of all-things-considered "oughts." When Mike has a very good reason for ignoring etiquette, then all-things-considered, he ought to ignore it, despite a remaining sense in which he should not

speak with his mouth full. Reasons go along with all-things-considered “oughts,” but not necessarily institutional “oughts.” The unusual thought of the person who stands to gain a million dollars by being rude shows that it may be the case that someone ought to ϕ without having a reason to ϕ . But this is not an all-things-considered sense. By comparison, it would do violence to our concepts to hold that a person ought, all-things-considered, to ϕ , but has no reason to ϕ .

Moral prescriptions are all-things-considered judgments, though not all all-things-considered judgments are moral. If I decide, after weighing all considerations, that I ought to pop down to the café at one o’clock rather than noon, it does not make the associated ought-statement a moral imperative. But that all moral judgments are all-things-considered judgments seems correct. The moral judgment “Mary ought not to have lied” has an overriding quality. It may be that according to prudence Mary ought to have lied, according to etiquette she ought to have lied, according to a utilitarian concern for the interests of the person addressed she ought to have lied, but if, morally speaking, she ought to tell the truth, then there is nothing else to consider when she asks herself what she ought to do. That morality prescribes ϕ -ing implies that all the other prescriptive systems have already been taken into account in so far as they are relevant, and balanced. In contrast, that etiquette may prescribe ϕ -ing does not imply anything about the requirements of other systems, and it does not imply that ϕ -ing is what she must do.⁶

A body of rules, according to Foot, yields prescriptions that we have reason to follow only if we have a standing interest in that body of rules. For most of us, etiquette does provide prescriptions that not only apply to us in the weak categorical sense, but which we must follow, because we care about etiquette. We have good reasons not to speak with our mouths full, but these reasons are escapable. Her complaint about a dominant thread of post-Kantian ethics is that morality is perceived as a body of rules that we have reason to follow regardless of our interests, desires, or concerns about morality.

We may consider a case of an unwilling gladiator. When the gladiator, against his will, is thrust into the arena, he immediately becomes subject of various rules of conduct: he must not throw sand in his opponent’s eyes, he must not drop his weapons and run, for instance. These are weak categorical imperatives. Nevertheless, the gladiator may have no reason to refrain from throwing sand in his opponent’s eyes. He may have no reason to do what he ought to do. Let us say that the unwilling gladiator has various imperatives *applying to* him whether he likes it or not, but it is not the case that he must *follow them* whether he likes it or not. If the authority of morality were no more than imperatives applying, then it would have no more authority than the Romans had over their unwilling prisoners. It would just be a set of rules which we have decided to apply to everyone, possibly with the threat of punishment for violators. But ordinary moral thinkers understand an unrepentant criminal differently from an unwilling gladiator. It is not merely that there exists a body of rules backed with punishment, and that the criminal has violated those rules. We think that he must follow those rules whether he accepts them or not. We think that the rules are, in some sense, *his* rules whether he accepts them or not. In trying to make sense of this inescapability, the most promising avenue is to seek a theory of reasons whereby the felon would have reasons for good behavior. Suppose he were to say, “I admit that I ought not to have broken my promise, but that is just an according-to-the-rules-of-morality-ought. I do not have an interest in that system of rules. Following that system of rules goes against my interests and desires. Therefore I have no reason to obey it.” Our moral discourse certainly does not fold in the

face of this. We insist that the felon, all-things-considered, ought to have kept his promise, and that he has a reason to follow the moral injunction. The question is whether these insistences are philosophically defensible.

One strategy for arguing for moral inescapability would be to find a normative system for which there is agreement that it fulfills the kind of inescapability we have considered, and then to show that moral imperatives are entailed by this system. By casting doubt on the entailment, Foot criticizes the popular strategy of securing moral inescapability via identifying moral prescriptions with rational prescriptions. All instances of this program, in her opinion, rest on some “illegitimate assumption.”⁷ “Irrational actions,” she says, “are those in which a man in some way defeats his own purposes, doing what is calculated to be disadvantageous or to frustrate his ends. Immorality does not *necessarily* involve any such thing.”⁸ That she focuses her criticism on the attempt to forge a necessary link between moral action and rational action suggests that rational imperatives may possess the kind of inescapability she denies to morality. In the interests of understanding inescapability, we should pause on this thought.

The mere possibility of being able to call a morally criminal action “irrational” is not sufficient, without further discussion, for concluding that we must have reason for refraining from performing it. If *R* is a rule of etiquette, then any violator may be called “rude,” but this is insufficient for making *R* inescapable. In such cases, the permissibility of using evaluatively-loaded language just reflects the fact that etiquette and rationality are normative systems, and in neither case is anything revealed about the agent’s reasons. Furthermore, there is cause for doubting the equivalence between an action’s being rational and there being a reason for performing an action. If I am lost in a mountainous forest, then, given my desire to escape, the rational thing for me to do is to head downhill, find a stream, and follow it. This remains so even if, unknown to me, a hundred yards uphill is a road and a gas station. The presence of the gas station provides a reason for me to go uphill, but does not make it rational for me to go uphill. An omniscient observer, seeing me to set off uphill, would conclude that I was acting in a thoroughly irrational manner.

But even if sentences of the form “*S* has a reason to ϕ ” and “It is rational for *S* to ϕ ” are not equivalent, it still may be that sentences of the form “It is rational for *S* to ϕ ” entail sentences of the form “*S* has a reason to ϕ ”. In the case of being lost, though I had reason for acting irrationally, I also had reason for acting rationally: even if heading off downhill would result in my quick demise at the paws of a grizzly bear, I still had a reason for going downhill. We may call the reason that I am completely ignorant of the objective reason. When I am justified in believing myself to have an objective reason, as on this occasion I am justified in believing myself to have an objective reason to go downhill, then I have a subjective reason. I am practically rational in so far as my actions are guided by my subjective reasons.⁹ It follows, then, that sentences of the form “It is rational for *S* to ϕ ” do entail sentences of the form “*S* has a reason to ϕ ”, allowing us to say that if it is rational for me to ϕ , then I have a reason to ϕ , whether I like it or not. If it is rational for *S* to ϕ , then *S* cannot just turn around and say something of the form “But I do not care about ϕ -ing; I have no reason to do so.” There is a reason for *S* to ϕ , just in virtue of what we mean by “rational” and “reason.” It does not follow, however, that the reason is inescapable. That depends on what theory of rationality we adhere to. If an action is rational only in so far as it furthers the desires of an agent, then we can alter what it is rational to do by altering our desire states. If such a Humean conception of practical rationality were correct, then it would not matter if

moral prescriptions were rational prescriptions, since all rational prescriptions would be escapable. Bearing this in mind, since Foot is a Humean about rationality, then we may wonder what kind of inescapability rationality is being implicitly accorded.

Although, for a Humean, any particular rational prescription is escapable, we cannot escape the system of practical rationality. Practical rationality, conceived of as a normative system yielding rules, is an inescapable set of escapable prescriptions. The worry, then, may be that the same cannot be said of the normative system of morality. Not only can any moral prescription be evaded, but, by altering desires and interests, we may have no reason to accept the moral system at all. Foot appears to think this true of any normative system except rationality. Why is rationality *sui generis* in this respect? The special feature of a theory of rationality is that it is a theory of what, in general, reasons are, when we have them and when we do not.

Humean instrumentalism may be presented as a general biconditional targeting having-a-reason for action: "For any x , iff x desires X , and ϕ -ing is the best means of achieving X , then x has a reason for ϕ -ing." The analysis may be turned into a conditional imperative: "If you desire X , and ϕ -ing is the best means of achieving X , then ϕ ." Can we say "I do not care about the conditional imperative; I will not follow it. I have idiosyncratic desires that are not furthered by following it. I have no reason to follow it"? We need imaginatively to locate an agent who really does not want to follow the conditional imperative, who in fact has no desires that are likely to be satisfied by following it, and then ask: "Ought she follow the conditional imperative? Does she have a reason to follow it?" But we cannot imagine such a being. Start with the assumption that an agent has unusual desires, such that following the conditional imperative does not conduce to the satisfaction of the desires. If the desires are such that following the conditional imperative will not tend to their satisfaction, then, if we assume that they are possibly satisfiable by the agent, we must conclude that there is something else that does promise to satisfy them. Doing something else will tend to the satisfaction of the desires of the agent, so the agent has no reason to follow the conditional imperative, but instead has a reason to do something else. But this undermines the initial assumption. The agent has reason to follow the conditional imperative after all, for it mediates the connection between doing something else and satisfying the desires.

Since, on the assumption that the conditional imperative is derived from the true theory of practical rationality, we cannot conceive of the example where the imperative "You ought to be practically rational" may be escaped via citing unusual desires, the imperative must have a kind of inescapability. This conclusion was not reached in the straightforward manner. It was not that we could imagine an agent who had desires that following the conditional imperative would not help satisfy but who had reason to follow it regardless; we found that we could not imagine an agent who had desires that following the conditional imperative would not help satisfy, because of the content of the conditional imperative. It is not a principle that practical rationality may or may not require, but constitutive of practical rationality.¹⁰

If this is the correct account of the kind of inescapability which practical rationality may be accorded, it is also an explanation of why practical rationality will be *sui generis* in this respect. Where, then, might moral inescapability stand? One possibility is that morality and practical rationality are identical, in which case, if we endorse a Humean instrumentalism, morality too will be an inescapable set of escapable prescriptions. Another possibility is that the general prescription "Be moral" is itself a particular imperative of practical rationality, in which case morality will be an escapable set of escapable prescriptions. Neither option is

satisfying to the defender of moral inescapability. Neither leaves her with capacity to censure the uncaring, conflict-free criminal.

An obvious strategy for a Humean with an interest in defending the inescapability thesis is to locate a universally held, perhaps necessarily held, desire or interest. Suppose there were some desire that all people have, and that the imperatives “Keep promises” and “Do not steal” were sound pieces of advice on how to satisfy that desire. Then we would all have reason to be moral, and in so far as the desire would be inescapable, so too would be the reason. Such inescapable imperatives would still be, in Kant’s terms, hypothetical, in so far as they remain dependent on our having a further end, but we all have that end. He calls them “assertorial practical principles.”¹¹

The prime contender for such a universally held desire is self-interest. The program, then, will be to show that moral actions, which importantly encompass seemingly selfless acts, can find a basis in enlightened selfishness. Up to a point this Hobbesian program is likely to meet with every success. It probably can be shown that most people in ordinary situations have solid, selfishly-based reasons for acting in accordance with morality. Famously, though, Hobbes makes little headway against the Lydian shepherd, whose selfishness it was that led him to steal, kill, and set himself up as king. We can allow that, given facts about the kind of creature he is, it is in the shepherd’s interest to have friends, lovers, family, and even to feel part of a cooperative community. But he may satisfy such desiderata, and do so in a sincere, non-coercive manner, while still using his ring to inflict great harm upon more distant others. To think that the Lydian shepherd, when a self-made king, must live a cold, unsatisfying existence, friendless and unloved, with troubled conscience and damaged soul, is simply silly.

While the Hobbesian program will not provide moral inescapability, it may provide reasons for ordinary persons in ordinary situations to act in accordance with moral imperatives. That may be a highly worthwhile pursuit. But if our moral discourse is exactly that with which we would hope to address the Lydian shepherd or the unrepentant criminal, then a Hobbesian does not end up vindicating morality. By comparison, showing that it is often a good idea to open a chess game with your king’s pawn does not suffice for making that opening an unbreakable rule of chess. Were we able to show that it is *always* the best opening, it still would be insufficient. If a system of imperatives leaves the harmful wielder of the ring of Gyges uncensured as he wreaks havoc then it is not a moral system. If the system does succeed in providing all ordinarily situated people with reasons to behave, that does not suffice to make the reasons moral.

As we have considered it, the failure of self-interested reasoning to prescribe in accordance with morality is a purely contingent matter. There is, however, in addition a conceptual concern, that moral imperatives are principles opposed to self-interested imperatives. W.K. Frankena, for example, suggests that “prudentialism or living by the principle of enlightened self-love just is not a kind of *morality*.”¹² In the Judaic-Christian tradition, morality is exactly that with which we attempt to counter selfishness. However, Frankena’s concern need not be discussed further, because the contingent failure of prudence and morality always to coincide is clear enough. We might instead look for a universally held non-selfish desire, such as a sense of sympathy, or fellowship. They will not serve, however, because a thesis of natural sympathy is only sensible when this sentiment is presented as a natural disposition or propensity. To present the-good-of-fellows as the object of a desire which all people have, a desire from which no one could escape, is to divest the thesis of

much of its attraction. We are all too familiar with counter-examples. Besides, we may credit the wielder of the ring of Gyges with all sorts of non-selfish desires. Perhaps his caring for the interests of his friends, family and community is not motivated by self-interest at all. But none of this will be sufficient for grounding imperatives proscribing his inflicting harm upon the inhabitants of the neighbouring valley.

Humean hopes of finding a universally held desire or interest upon which moral inescapability may be grounded are bleak. The one credible candidate for being universal is simply not going to provide the correct content all of the time. Of course, it is not necessary for a Humean to find a universal desire that being moral will serve. It would suffice that we all have some desire or other that being moral will serve. This clarification, however, will not further the cause. There is nothing plausible in the claim that, regardless of how a person is situated, he will have desires and interests that promise-breaking and stealing will frustrate.

That moral inescapability cannot be squeezed from a Humean framework is hardly surprising. Hume and Foot do not wholeheartedly attempt the operation, and are content if most people, most of the time, have reason to act in accordance with morality. They acknowledge that they cannot definitively speak to the bearer of the ring of Gyges. That this appears to be such an unpopular position reveals a feeling that there being such contingent limits on people's reasons does not capture what we mean by "morally ought" at all. This is correct. By "morally ought" we mean something much stronger. Our moral discourse, if employed in a Humean world, systematically fails. One way of establishing this failure would be to set out to defend Humean practical rationality, or instrumentalism. One way of avoiding this failure would be to defend a non-Humean rationality, or non-instrumentalism.

How might instrumentalism about reasons be established? Bernard Williams attempts to show that the opposition necessarily commits an error.¹³ Reasons, he argues, are items that can be appealed to as potential explanations for the actions of agents; only items that are connected in a certain way to an agent's motivational set can play such a role; a non-instrumentalist view of reasons entails the existence of reasons that could not play such an explanatory role; therefore non-instrumentalism is false. Non-instrumentalism implies the existence of what Williams calls external reasons, but, he argues, there are no such things. If Williams's argument is coupled with the plausible premise that moral discourse is committed to the existence of true external reason statements, then we have an argument for an error theory. This argument, however, has an inevitable weakness. Why should we believe Williams's initial claim about what reasons are? We are, as philosophical inquirers, unsure what reasons are, and we are attempting to explicate the concept in an adequate way. Williams offers an explication which is, at first blush, credible enough, but before we endorse it we must think through its implications. Does it have consequences that clash with any firm intuitions that we have? This is where the difficulty in using Williams's argument as a premise for a moral error theory arises. It is a dominant intuition that everybody has a reason to refrain from, say, killing innocents for fun, regardless of desire or circumstance, and if a theory of reasons fails to underwrite that intuition, so much the worse for that theory of reasons. In general terms, we cannot take a controversial concept like having-a-reason, explicate it in a certain way, and then show that the explication implies the falsity of a widely held, intuitively attractive discourse. The fact that it has such an implication will simply be taken as evidence against the explication. This is how Michael Woods argues about practical reasoning: Humean instrumentalism cannot be the correct account of practical reasons

because it cannot account for moral or aesthetic reasons, which are, sometimes at least, non-instrumental.¹⁴

This raises the concern that moral discourse may be methodologically immune from systematic failure, and that any philosophical thesis having a moral error theory as a consequence may be rejected simply in virtue of possessing such an outrageous consequence. This is of concern not because a moral error theory is a particularly desirable conclusion, but because it ought at least to be recognized as a real possibility. Although a moral error theorist may, at the end of the day, be defeated, it should not be the case that anything that counts in her favor may automatically be discounted just in virtue of counting in her favor. If her moral error theory were clearly and obviously false, that would be another matter. But that nothing so clear and obvious is the case should be apparent from the fact that we have always had a cluster of vague hypotheses about the genealogy and function of moral language according to which the whole discourse is, to put it most bluntly, bluff.¹⁵ This hypothesis is certainly not well-liked, and may well be false, but it is certainly coherent, and therefore should be taken seriously. If we set up the rules of the philosophical game such that the hypothesis cannot leave the starting blocks, then so much the worse for that rule book.

There are many ways in which the hypothesis might be presented. We could tell a genealogical story which highlights the supposition that moral discourse grew up entwined in theistic discourse: human beings were conceived of as God's creatures, and our rules of conduct were simply his laws. We may then, for any number of familiar reasons, decide that the theistic picture of the world is utterly mistaken. This may lead us to speculate about why we adopted the theistic moral viewpoint in the first place, and an obvious train of thought would have us see it, at least in part, as a tool of political and interpersonal manipulation. On this pessimistic view, accepted moral obligations are just the rigidified rules of a community: certain behaviors are socially expedient and they become cloaked in a pseudo-mystical aura of inescapability. At one time the inescapability was cashed out in explicitly supernatural terms, and now, when that framework is widely rejected, or at least uncoupled from morality, the social need for a sense of inescapability remains. Inescapability, on this view, might cynically be described as a manipulative fiction. The cynicism pertains not to the fictional but to the manipulative part of the description. Fictions, after all, may well be useful for the smooth running of the community, and there need be nothing intentionally deceptive about them.

C.L. Stevenson presented a cool-headed version of this view.¹⁶ According to his early account, a moral judgment, such as "Promise-breaking is wrong," amounts to an assertion about ourselves plus a command: "I disapprove of promise-breaking; do so as well." As a theory of the meaning of moral language, Stevenson's view is unconvincing for a number of reasons.¹⁷ As a theory of how moral language is used, it has more merit.¹⁸ It is uncontroversial that sometimes moral language is used largely in order to secure the agreement of others. We are all familiar with poorly conducted, heated debates, in which the moral concepts play the role of rhetorical heavy artillery. This does not imply that the purposes of the interlocutors are transparent to themselves. We may credit them with sincere beliefs that something is morally right or morally wrong, but the beliefs are merely rigidifications or rationalizations of pre-existing opinions. We are familiar enough with the thought that on certain occasions with certain persons, moral debate is essentially bluff. If we can make sense of it on some occasions, then we can make sense of it in general. It may well

be that all moral debate has this status. The bluff hypothesis is just a well-known appraisal of certain shabby moral disputations, writ large.

It is important to see how it might be that inescapability is a fiction. That we have a popular discourse which embodies the notion of inescapability provides some *prima facie* favoring of a theory that purports to explicate and vindicate it. There is therefore *prima facie* weight against any piece of philosophical reasoning that has as an implication that the discourse is hopelessly flawed. That is why the strategy of an error theorist of giving an account of having-a-reason in such a way that a moral error theory will be a natural conclusion is vulnerable. When we proffer philosophical explications, intuitions play a substantial role. For all that, the counter-intuitive account of having-a-reason may well be correct, but to see this it will be important that it be supplemented with a plausible account that casts the relevant intuitions into doubt. This is the role of the bluff hypothesis, rough and speculative as it is. It is not merely a claim that something that we typically believe in is a fiction. If told in a certain way, the hypothesis suggests an account of moral phenomenology, and therefore predicts that we will find intuitively attractive something that is false.

Consider, for example, a projectivist account of our moral lives. Certain events, states of affairs, and character traits, prompt a certain kind of subject, in certain circumstances, to have a certain kind of psychological response. The subject then projects the response onto the event, and thus “raises, in a manner, a new creation.”¹⁹ This is the Humean picture of the mind’s “great propensity to spread itself on external objects.”²⁰ If an act of stealing makes me angry, I project this sentiment onto the world and see that act as having the property of anger-demandingness, must-not-be-doneness, or simply wrongness. In this way the projectivist picture includes an account of how moral properties appear to us. They appear as if they are in the world. By the hypothesis of projectivism, moral realism is intuitively attractive. *Ex hypothesi*, projectivism is counter-intuitive. Intuitions in favor of moral realism, thus, do not tell against projectivism, since projectivism would have us predict that such intuitions will be in place.

There will be different ways of presenting the bluff hypothesis, but what they will have in common is that judgments that certain imperatives to action are inescapable are never true and are beyond the pale of philosophical defensibility. The notion of inescapability is a well-worn practical tool with which we manipulate and negotiate the social world, often successfully. Such a hypothesis allows us to speculate about the function of appeals to inescapability. It suggests why we would feel the need to insist upon reasons when there are none; it suggests that we would find a fiction beguiling. Therefore the knee-jerk intuitions against the hypothesis carry no weight. None of this is yet to claim that the hypothesis is credible or to be preferred to the opposition.

We have not yet seen any reason for accepting a bluff hypothesis. It has only been suggested that the fact that it ill-fits our ordinary thoughts and intuitions may be discounted as evidence against it. The plausibility of any precise version will depend on its details. The most telling point in favor of the general view that appealing to inescapability amounts to bluffing is a comparison with the alternatives. The important alternative is that it is true that persons must act in certain ways whether or not it furthers their interests or desires, and that they must have associated external reasons for so-acting. But we do not have a credible, non-dogmatic theory that makes sense of this thought. We find ourselves with a strong desire to say that a moral felon, despite his lack of repentance, his ring of Gyges, and the fact that he knows himself to be terminally ill, ought not to break a promise or a stranger’s arm. We have

a strong desire to distinguish the position of such a criminal from that of an unwilling gladiator. Yet when asked to elucidate this claim we immediately founder. Centuries of philosophical foundering over a familiar, seemingly straightforward thought should prompt suspicion. A theory that not only would have us assert that there is no sense to be made of the thought, but also explains why we would typically make the error, gains support.

Unsurprisingly, we may also find support for the conclusion in the work of John Mackie in his “argument from relativity.”²¹ The argument starts from the observation that there are various incompatible moral systems held by different cultures and different individuals. Although it might be true that one group has it right and all the others are wrong, a better explanation is that nobody has it right because there is nothing to get right. The preferred explanation is that moral judgments are reflections of cultural placement; they are inventions and not discoveries. Similar considerations may tell against religious beliefs. There are many incompatible religious systems in the world. If I am a believer in the system that my parents or my culture handed me in childhood, then I must think that many of the other religions are mistaken. I will also be inclined to form an opinion about how and why all the other false religions developed. It is likely that I will interpret them as practically useful fictions: the communities in question invented their Gods and their religious injunctions in order to satisfy a variety of pragmatic needs. I may be inclined to go so far as to read their religious discourse as bluff in the service of social and political manipulation. I will not be content with my own community’s religion until I have an acceptable theory about how and why the other religions are expedient fictions. But its very credibility ought to cause me to worry. If religion-*qua*-expedient-fiction is the most likely hypothesis for all the other communities, why is it not also for my community’s religion? If I find it plausible there, why should I not find it plausible here?

Mackie’s argument from relativity about morality may be spun in a similar way, and we can do the same for inescapability. We all observe many seemingly intractable debates concerning what people must do. It is clear that not all participants can be correct. If I am inclined to endorse one side, then I must see the opposition as being in error, and if I am thoughtful, I will want to know what led them to that error. A common, highly plausible assessment will be that the claims of the opposition are self-serving rationalizations of idiosyncratic opinion. They have a strong feeling in favor of one course because there is a lot at stake for them, or because they were raised in a certain way, and an insistence that a certain course of action is required or forbidden is the strongest rhetoric available to them for promoting that end. But if I have any imaginative capacity to detach myself from my viewpoint, I may come to see that if this is the best explanation of my opponent’s insistence, then it may well be the best explanation of my own inclination to insist. Unless I can find some distinguishing feature of my own viewpoint, then I ought not prefer it to my opponent’s. This, as we have seen, is how the bluff hypothesis may operate. If we find it the best explanation of some debates, then perhaps it is the best explanation for all debates, including moral discourse in general. Naturally, we do not find that a welcome conclusion, but that is exactly what a proponent of the hypothesis would predict. When we interpret a particular moral contention as heartfelt sham, we do not thereby expect that the brandisher of the judgment will also be inclined to agree with us. On the contrary, bluffers do not relish their bluff being called.

The only incentive we would have for resisting an explication of having-a-reason like that of Williams is if we foresaw that it threatened the thesis of moral inescapability, and did not

welcome that. As a claim about philosophical intentions, the observation is merely *ad hominem*, and perhaps has no place here. As a careful claim about what may be counter-intuitive about Humean instrumentalism, the observation is more serious. If we have a hypothesis according to which persons would naturally find statements of external reasons plausible and compelling even though false, then we have something with which to oppose any appeal to pre-theoretical intuition from a non-instrumentalist. If, in addition, we note that she has no clear, credible account of non-instrumentalism, then we have occasion for positive suspicion. If we can muster arguments that provide some confirmation for the hypothesis that we would have reasons for making false reason statements, then we have grounds for taking a stand. That, perhaps, is the best that can be done. Were a certain popular discourse to revolve around defective concepts, then, short of showing those concepts to be self-contradictory, it would always be difficult to prove their defectiveness. We can point out that there is little sense to be made of them, but adherents of the discourse will always promise the solution in the near future. Naturally, when we think that someone has gone missing in the wilderness, we should always send out a search party. But if, after a good long search, nobody is found, then we should consider the possibility that we were mistaken in thinking anyone was lost in the first place. If there are some philosophers who wish to keep seeking the fugitive indefinitely, there is little to be said to them.

But if inescapability is at bottom just a crazy idea, we might well wonder just what we are thinking when we appeal to it. In his "Lecture on Ethics," Wittgenstein opined that our moral language seems dominated by similes.²² Mackie's talk of obligation being an "invisible cord" and a demand for payment being an "immaterial suction-pipe" dredging for the owed money, comes to mind.²³ Mackie's images are attempts to catch the feel of inescapability, and perhaps all we need is a shared feel in order to communicate successfully. When Wittgenstein looks behind the apparent similes for non-metaphorical elucidation, he finds nothing at all, and concludes that moral discourse is "mere nonsense."²⁴

The discovery that morality is nonsense does not incline Wittgenstein to advocate its abandonment. Though the "very essence" of morality is its nonsensicality, it is something he "cannot help respecting deeply" and he refuses to ridicule or "belittle this human tendency."²⁵ There is no reason to doubt that his day-to-day appeals to nonsense served Wittgenstein well. Appealing to inescapability is often useful. Indeed, part of the argument for preferring the view that there is no moral inescapability is that we would find it pragmatically expedient to invent it even if it did not exist. Moral discourse should not be eliminated. Even if we know that we have no business telling the Lydian shepherd or an unrepentant criminal that he ought not to have done what he did, when in the thick of social interaction such utterances and such thoughts may be not only convenient, but of considerable practical advantage. It does not follow that we should continue with an array of false beliefs on the matter, because adopting the moral point of view need not involve belief. Allowing the thought of inescapability to reside in the forefront of our mind, as a controlling image or regulative feeling, may be entirely sufficient for us to garner whatever utility moral discourse typically brings. So understood, the recommendation "Adopt the moral point of view" may be sound practical advice, without the pretense of being universal or inescapable.

Notes:

1. Philippa Foot, "Morality as a System of Hypothetical Imperatives," *Philosophical Review* 81 (1972), p. 311.
2. John Tilley, "Motivation and Practical Reasons," *Erkenntnis* 47 (1997), p. 121.
3. David Hume, *Enquiry Concerning the Principles of Morals*, (Cambridge, Mass.: Hackett Publishing Company, 1983), p. 82.
4. See Plato, *Republic*, Paul Shorey, trans., (London: William Heinemann Ltd., 1930), p. 117 ff.
5. Gilbert Harman connects the predictive "ought" to reasons by claiming that "There ought to be an avalanche soon" may be analyzed as "There is good reason for believing that there will be an avalanche soon". See Gilbert Harman, "Reasons," *Critica* 7 (1975), and John Mackie, *Ethics: Inventing Right and Wrong*, (New York: Penguin Books, 1977), p. 74.
6. See Lawrence Becker, "The Finality of Moral Judgments: A Reply to Mrs Foot," *Philosophical Review* 82 (1973).
7. Foot, *op. cit.*, p. 310.
8. *Ibid.*, p. 310.
9. See Garrett Cullity and Berys Gaut's introduction to *Ethics and Practical Reason*, G. Cullity and B. Gaut, eds., (Oxford: Clarendon Press, 1997), p. 2.
10. See James Dreier, "Humean Doubts about the Practical Justification of Morality," in G. Cullity and B. Gaut, eds., *op. cit.*
11. Immanuel Kant, *Groundwork for the Metaphysics of Morals*, H.J. Paton, trans., (London: Routledge, 1989), p. 78.
12. W.K. Frankena, *Ethics*, (New Jersey: Prentice-Hall Inc., 1963), p. 18.
13. See Bernard Williams, "Internal and External Reasons," in his *Moral Luck*, (Cambridge, England: Cambridge University Press, 1981), and "Internal Reasons and the Obscurity of Blame," *Logos* 10 (1989).
14. See Michael Woods, "Reasons for Action and Desires," *Proceedings of the Aristotelian Society*, Supp. Vol. XLVI (1972).
15. The term "bluff" in this context is derived from Williams's choice of the word, in Williams, (1981) *op. cit.*, p. 111.
16. See C. L. Stevenson, "The Emotive Meaning of Ethical Terms," *Mind* (1937), and *Ethics and Language*, (New Haven, Conn.: Yale University Press, 1944).
17. See Peter Geach, "Assertion," *Philosophical Review* 74 (1965).
18. See Alasdair MacIntyre, *After Virtue*, (Notre Dame: University of Notre Dame Press, 1984), Ch. 2.
19. Hume, *op. cit.*, p. 88.
20. David Hume, *A Treatise on Human Nature*, L.A. Selby-Bigge, ed., (Oxford: Clarendon Press, 1978), p. 167.

21. Mackie, *op. cit.*, pp. 36-8.
22. See Ludwig Wittgenstein, "Lecture on Ethics," in Stephen Darwall, Allan Gibbard, Peter Railton, eds., *Moral Discourse and Practice*, (New York: Oxford University Press, 1997).
23. Mackie, *op. cit.*, p. 74.
24. Wittgenstein, *op. cit.*, p. 69.
25. *Ibid.*, p. 70.