

Normative Explanations

by

Geoffrey Sayre-McCord

University of North Carolina at Chapel Hill

Introduction

*there is a great wind of moral force moving through the world
and every man who opposes that wind will go down in
disgrace*

So wrote a widely respected historian, political scientist, and United States President -- Thomas Woodrow Wilson.¹ Nowadays, Wilson's confidence in morality, his conviction that justice will out, that the righteous will rule, that morality can and does have an impact on what happens in the world, seems anachronistic, outrageously optimistic, and metaphysically peculiar. That it is anachronistic and outrageously optimistic, I don't deny. But I do hope to show that it is not nearly as metaphysically peculiar as many think. In particular, I hope to make sense of the idea that morality operates, if not as an over-powering wind, then at least as a gentle breeze in the course of history.

Less metaphorically, I want to explore the view that morality, and specifically moral rules, might help to explain social change.² I should emphasize, though, that my aim is to *make sense of the idea*, not to show that in fact morality actually does explain social change; I am after a plausibility argument, not an existence proof. Providing the first is hard enough.

Before moving directly to *moral* rules, and their potential role in explanations of social change, I will concentrate first on the more general issue of how rules (whether moral, legal, or social) figure in explanations. In what follows, I will be moving back and forth frequently between moral rules and other (metaphysically less presupposing) rules. Throughout, though, I shall be

concentrating on normative rules -- rules that forbid courses of action, impose obligations, or grant permissions -- not on descriptive rules that serve simply to express empirical regularities. The contrast, I think, is intuitively clear, even if it doesn't lend itself to a rigorous definition: it is one thing to say that, in the U.S., there's a rule (in this case a law) forbidding speeds in excess of 65 mph; it is quite another to say that, as a rule, people in the United State don't drive over 65 mph. Likewise, it is one thing to say that one ought to make rational choices; it is another to say that people do, as a rule, make such choices. And again, it is one thing to say that in baseball one is permitted by the rules of the game to steal second base; it is another to say that, as a rule, people do steal second base.³ Mere regularities do not constitute normative rules, even though recognized and enforced normative rules may be expected to give rise to regularities. Whether normative rules that are unrecognized, or unenforced, or both, also give rise to regularities is less clear. That they do, or at least might, is something I argue for later in the paper.⁴

Normative rules are, it seems, ubiquitous; they apparently structure our interactions, inform our plans, define our options, and play a central role in our understanding of our own activities.⁵ Yet the importance of normative rules, when it comes to explanation, as opposed to exhortation, remains dubious, at best.

Of course normative rules might explain why some action, for instance, counts as illegal, some utterance as ungrammatical, some proposal as rational, some institution as moral. Left lingering by this observation (right as it might be) is the suspicion that the facts explained by the relevant rules are just reflections of the rules. To say an act is illegal is, one worries, simply to say it violates a law; to say an utterance is ungrammatical is simply to say it violates a rule of grammar, and so on. Much more satisfying would be the discovery that the normative rules and corresponding normative facts they 'explain' themselves explain -- *really* explain -- some event, process, or situation, that is conceptually

¹. Quoted in John Morton Blum's Woodrow Wilson (Boston: Little, Brown, & Co., 1956), p. 159.

². Of course moral rules constitute neither the whole nor even the major part of morality. I concentrate on *rules* for two reasons. The first is that moral rules appear to be disembodied in a way that makes them especially problematic. Thinking that the rules of justice might explain social change is, initially at least, much less plausible than thinking, for instance, that someone's courageousness might explain why she does what she does. The second reason is occasion specific: the conference for which this paper was written revolved around the proposal that a logic of rules might contribute to explanations of social change. Two convictions guided the proposal: (i) that the social rules the logic could be used to represent themselves play a role in social change and (ii) that the rules of the logic could in turn help to explain why the social rules play the role they do. Both convictions presuppose that normative rules (whether normative social rules or normative rules of logic) can explain things that happen in the world.

³. Many (though not all) of the normative rules I have in mind contrast as well with the sort of 'ought' judgments that resolve neatly and easily into empirical 'if-then' claims. For instance, unlike "the gasoline ought to have an octane rating of at least 92," which might well be cashed-out as "if the gasoline has a rating of less than 92, then the engine won't run without pinging," claims like "You ought (as a matter of law) to drive under 65 mph" are less plausibly translated into 'if-then' substitutes that avoid mention of the rules in question. Others, though, for instance the requirements of practical rationality, seem at least to be candidates for such a reduction (e.g. if you don't do *x*, then your preferences will likely be more frustrated than they otherwise would be).

⁴. Even if, as I believe, there are some unrecognized and unenforced normative rules that apply to people and societies, it seems one could have no reason to ascribe a rule to a society *as a rule of that society* unless it was at least implicitly recognized and enforced. For a discussion of this point, see Philosophy of Social Science, David Braybrooke (Englewood Cliffs: Prentice-Hall, 1987), pp. 47-57.

⁵. See The Reason of Rules, Geoffrey Brennan and James Buchanan (Cambridge: Cambridge University Press, 1985).

independent of the rules in question.⁶ The nagging suspicion is that normative rules explain no such thing.

The doubts about the explanatory importance of normative rules come from many directions. Perhaps the most influential has its source in the conviction that there is a fundamental difference between fact and value, a conviction bolstered by the observation that what ought to be isn't always so, and that when it is so, it apparently isn't so *because* it ought to be. The inclination is to argue that only facts *as distinct from values* explain things that happen in the world. Notoriously, though, the distinction between facts and values is extremely difficult to articulate. Moreover, even those convinced of its existence are usually willing to count many normative social rules (such as those governing language, constituting legal systems, and defining games), even if not moral rules, as firmly on the side of facts.

Two Ways to Figure in Explanations of Social Change

If normative rules do figure in explanations it will be in either of two roles: as part of what is *being explained* or as part of what *does the explaining*. In the first case, when it comes to explaining rules, we would be interested in explaining the origin and stability of the rules in question and in explaining the various changes they might undergo. In the second case, when it comes to explaining by appeal to rules, we would be interested in using rules to help explain other rules, to explain particular actions, or to explain general trends and patterns. Either role, that of explainer or that of explained, will give normative rules a place in explanations.

Still, the two roles are not of equal significance. Normative rules only become interesting objects of explanation to the extent they are also themselves explainers -- to the extent the rules being what they are makes some difference to what happens in the world. If rules never make such a difference, and so never help to explain what happens, we would have much less reason to bother explaining them. In fact, it is hard to imagine what grounds one might have for believing the rules exist if they have no impact.⁷ Explained non-explainers seem just epiphenomena by another name with all the attendant problems. Nonetheless, before turning to how normative rules might explain, I would like

⁶ It might be satisfying enough if normative rules could be shown to explain normative facts that are conceptually independent of the rules doing the explaining. Many, though, would hold that normative rules are doing real explanatory work only if they contribute to explanations of non-normative facts, for instance only if they helped to explain some of our perceptual experiences. Behind this more stringent requirement is the conviction that an adequate epistemology must ground justification in sensation. See my "Moral Theory and Explanatory Impotence," *Midwest Studies XII* (Minneapolis: University of Minnesota Press, 1988), pp. 433-457, and Gilbert Harman's *Thought* (Princeton: Princeton University Press, 1973).

⁷ The underlying assumption here is that we have reason to believe only in those entities, properties, laws, and rules, that contribute to our best explanation of our experiences. See my "Moral Theory and Explanatory Impotence," *op. cit.*

first to mention briefly a couple of the ways they might be explained. I do this in part because some of our best explanations of rules themselves appeal to normative rules.

Normative rules fall fairly neatly into two groups. The majority owe their existence and their force to various social practices and conventions; the rules of chess, the requirements of etiquette, the standards of grammar, and the laws of the land (to take a few examples) all come into being, and lay a claim on us, only because of things people do or have done. Such rules wouldn't be, and wouldn't matter, save for the particular local practices of people. These rules are reasonably thought of as conventional. Other rules apparently transcend convention; the demands of morality, the rules of logic, and the standards of practical rationality, all appear not to depend (at least in any straight-forward way) on what people do or have done. The former are unmistakably human products, while the latter hover above, and seemingly free from the influence of, human practices and conventions. Whether this distinction between conventional and trans-conventional rules marks a real difference in kind, or just a difference in the degree to which the rules depend on more or less local practices and conventions, is both unclear and controversial. One might, and people do, argue that the apparently trans-conventional rules of morality, logic, and rationality, are, despite appearances, conventional rules that find their origin and force in very widely shared, and perhaps socially necessary, practices or conventions.⁸

Regardless, to the extent the normative rules in question are dependent for their existence and force on what people do or choose, or on practices and conventions, they lend themselves to psychological and social explanation. Among the most elegant of such explanations are those that, by making use of decision and game theory, explain the rules as products of rational behavior.

Some rules, for instance laws handed down from 'on-high' by a dictator, are plausibly explained as rational solutions to problems of parametric choice (that is, as rational solutions to problems where the agent chooses unilaterally from among what is viewed as a fixed set of options). Thus we might explain a Dictator's imposition of a curfew as his least-cost solution to the problem of containing insurgency; given his preferences, situation, and nearly unchecked power, passing laws that constrain the liberty of others might be (in some sense) the rational thing for him to do. On a less grand scale, the rules parents establish for their children are often designed as unilaterally imposed rational solutions to problems that would otherwise arise.

Other rules we might explain on the grounds that they are rational solutions, not to problems of parametric choice, but to problems of strategic choice (that is, as rational solutions to problems where the agents choose interactively among

⁸ I sketch a version of (what might be called) 'conventional moral realism' in "Coherence and Models for Moral Theorizing," *Pacific Philosophical Quarterly* 6 (1985), pp. 170-190.

options that depend in part on the choices of others).⁹ Plainly, social rules do frequently solve problems of strategic choice; they solve problems of coordination and problems of competition (usually with the help of enforcement mechanisms), and most often problems that mix elements of competition with the need for coordination. Moreover, that they solve such problems is plausibly taken to be the reason they exist. Traffic laws, to use a popular example, seem largely to have come about in order to coordinate behavior in a context where it doesn't much matter to people what exactly the laws are, just so long as there are some. Criminal laws serve primarily, although not exclusively, as a way of resolving predictable conflicts among the interests of people within a society. And anti-pollution laws have been introduced explicitly to militate against the tragedy of the commons, which is due to people exploiting unmercifully what is held in common and protected by no one.

Recognized, respected, and enforced social rules time and again shift expectations, circumstances, and motivations, so as to ensure that people won't act in ways that make all (or at least most) worse off. Without them social life would be impossible. Game theoretic explanations recommend themselves as well, though, for social rules of a less salutary cast. In South Africa, for instance, the laws of Apartheid were quite clearly the instruments of class interest, explainable in part as an immoral solution to a problem of collective choice faced by those in power.¹⁰ In the same vein, a much more benign example can be found in casino gambling rules, which are calculated to give the 'house' a distinct advantage.

Still other rules may not themselves be consciously introduced solutions to choice problems, but instead by-products of social practices and conditions. Rules of etiquette, for instance, appear often to impose capricious constraints that are not themselves rational solutions to choice problems at all, despite their being predictable consequences of social conditions. Even in these cases, invisible foot explanations (that account for the rules as the unintended and detrimental consequences of rational behavior) and invisible hand explanations (that account for the rules as the unintended but beneficial consequences of rational behavior)¹¹ will sometimes be available. If they are available, such

⁹. See Jon Elster's *Ulysses and the Sirens* (Cambridge: Cambridge University Press, 1979) for a discussion of the difference between parametric and strategic choice.

¹⁰. I've concentrated in the examples on laws, but decision and game theoretic explanations seem to work too for many other normative rules, for instance those governing membership in trade unions, and the authority of umpires in baseball, as well as less institutionally enshrined rules of social organization. David Lewis' *Convention* (Cambridge, MA: Harvard University Press, 1969) sets out elegantly the general structure of game theoretic explanations for social conventions (among which are all sorts of conventionally established normative rules). See also Russell Hardin's *Collective Action* (Baltimore: Johns Hopkins University Press, 1982).

¹¹. Each person "intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of

explanations will make use of decision and game theory to explain the behavior, practices, and social conditions, that produce the social rules as by-products.

Importantly, the fact that some rule would be a rational solution to a problem that would exist if there weren't such a rule, can't explain the rule's existence unless the relevant people acted *because* the rule offered a rational solution. Of course, the people needn't have acted because they *recognize* the rule offers such a solution; they may embrace and enforce a rule because the rule offers such a solution without knowing that is why they act as they do. Yet explanations that rely on decision and game theory will go through as *explanations*, and not simply as justifications or just-so stories, only if the people involved are suitably rational -- only if their behavior is sensitive to what is rational under the circumstances. That people do live up to the demands of rationality, and act because of them, is a crucial presupposition of decision and game theoretic explanations.

Just exactly what the demands of rationality are, needless to say, is a matter of extensive debate even among those who take advantage of the formal apparatus of decision and game theory, along with the essentially maximizing conception of rationality that underlies it. Neoclassical economists, for instance, embrace an unabashedly narrow, even cynical, view that ties rationality to self-interest; while others hold that the preferences of rational agents may range over the welfare of others. And some hold that rationality is simply a matter of maximizing expected satisfaction of (brute) subjective preferences; others impose restrictions (usually having to do with reflection or information) on the subjective preferences that are to count; while still others replace subjective preferences in the account with objective interests.¹² Despite these differences, the common framework offered by decision and game theory provides a strikingly powerful explanatory tool.

Such explanations are often among the best we have of why many familiar rules and sets of rules have been imposed or adopted. Indeed, a powerful, though not always successful, strategy for figuring out why certain social rules have come into existence lies in discovering whose interests are served by the rules; where interests are served we are likely to find agents active. Clearly this works only as a heuristic strategy. It would be a mistake to claim that the mere fact that interests are served can itself explain the existence of rules; there must

society more effectually than when he really intends to promote it." Adam Smith, *The Wealth of Nations*, (New York: P. F. Collier & Son, 1905), bk. 4, chapter 2, pp. 160-161.

¹². See R. D. Luce and Howard Raiffa, *Games and Decisions*, (New York: John Wiley & Sons, 1957); Alan Hamlin, *Ethics, Economics and the State*, (New York: St. Martin's Press, 1986); David Gauthier, *Morals By Agreement*, (Oxford: Oxford University Press, 1986); Richard Brandt, *A Theory of the Good and the Right*, (Oxford: Oxford University Press, 1979); R. B. Perry, *General Theory of Value* (New York: Longmans, Green, and Co., 1926); Stephen Darwall, *Impartial Reason*, (Ithaca: Cornell University Press, 1983); and Peter Railton, "Moral Realism," *Philosophical Review* XCV (1986), pp. 163-207.

be some reason to think either that the people were aware of the benefits to come or that some feedback loop would have adjusted the rules had the interests not been served.¹³ When the strategy fails, when the rules in question are not plausibly explained as rational solutions to choice problems, it will either be because the rules simply are not rational solutions or, if they are, be because their being rational solutions doesn't (under the circumstances) explain their coming into existence -- say when the people don't recognize their interests, or when they haven't for some other reason lived up to the standards of rationality.

What is interesting, here, about decision and game theoretic explanations of conventional rules (whether of the rules directly or of the social practices that generate the rules as by-products) is not that they exhaust the sort of explanations we have for social rules -- they certainly don't -- but that they evidently explain social rules by appeal to the rationality of agents and thus to normative rules of rationality. Of course, the explanations will count as *normative* explanations only among those who accept the underlying theory of rationality as at least in the running as a normative theory of rationality. But this is a very large group. So let's turn from explaining rules to using rules to explain.

Explaining by Conventional Rules

To the extent we try to use normative rules to explain suitably independent events, processes, or situations, there are three ways the rules might enter into an explanation. In setting these ways out, I will concentrate first on conventional normative rules, only afterwards turning to the more problematic case of trans-conventional rules.

Most obviously, and least controversially, normative rules can figure in explanations as the *content* of agents' beliefs. People quite clearly often do things (say, drive under 65 mph or avoid splitting infinitives) because they believe doing so is required by certain rules (e.g. of law or grammar). Less often, perhaps, but no less clearly, people do things -- out of perversity, indignation, or independence -- because they believe doing so violates certain rules.¹⁴

It is tempting to stop here, saying that all the explanatory work done supposedly by rules is really done by appeal to beliefs, not rules. The beliefs do, of course, have rules as their content, but that hardly gives the rules themselves a

robust role in explanations. After all, some people do things because they believe in Santa Claus, but we can explain their actions without sharing their conviction. Santa Claus himself, as opposed to beliefs about Santa Claus, figures nowhere in our explanations of things that happen in the world. In the same way, people may do things because they believe there are normative rules, but we might still be able to explain their actions without sharing their conviction. Normative rules themselves, as opposed to beliefs about such rules, might figure nowhere in our explanations of things that happen in the world. No doubter of the explanatory importance of rules need feel any discomfort allowing rules this role.

Nevertheless, the temptation to view the explanatory role of normative rules so narrowly should dissipate once attention is turned to explaining those beliefs that have rules as their content. For in many cases at least part of the explanation of why people believe the rules require them to act in certain ways is that the rules do. Part of the explanation will involve there actually being the rules about which the people in question have beliefs. Often enough, if the rules were different, people's beliefs, and their behavior, would have been different as well. For example, for several years, and until recently, the national speed limit in the United States was 55 mph. When the limit changed to 65 mph people's beliefs about the limit changed too, as did their driving habits. Were the limit to change again it is reasonable to think people's beliefs and behavior would change once more. In short, people's beliefs about the speed limit are explained (at least in part) by the speed limit itself, by the normative rule that forbids speeds in excess of 65 mph. Obviously, people's beliefs are not always sensitive to what the rules actually are; sometimes people believe rules require something of them when the rules don't. In these cases something other than the rule will have to explain their belief. Even so, rules will often explain the beliefs people have about rules and so they will explain indirectly the behavior caused by those beliefs. When they do, normative rules will figure in explanations not just as the content of beliefs but also as causes of belief and behavior. Presumably, when it comes to explaining our own beliefs about rules we are committed to thinking the rules being what we take them to be is part of why we hold the beliefs we do.

Normative rules can figure as well, though, in explanations where their effects on a person's behavior are unmediated by her beliefs about the rules. Sticking with the example of traffic laws, we might explain why a particular person is driving within the speed limit by noting that she is disposed to drive at roughly the same speed as those around her, and that those around her have slowed to 55 mph because they noticed what she did not -- a new speed limit sign. Ignorant though she is of the new speed limit on that stretch of road, the fact that it is the limit helps to explain her behavior because it helps to explain the behavior of those around her (whose behavior in turn explains hers).

Switching from rule-following to rule-breaking behavior, it's clear as well that rules that are unrecognized, yet violated, by a person can contribute to an explanation of what happens (just as rules that are unrecognized but followed

¹³. See Jon Elster's *Making Sense of Marx* (Cambridge: Cambridge University Press, 1985), and his *Ulysses and the Sirens*, *op. cit.*, and Richard Miller's *Fact and Method* (Princeton: Princeton University Press, 1987).

¹⁴. While these claims are, I take it, quite uncontroversial, they have (not surprisingly) been controverted, standardly on the grounds that talk of beliefs (not just beliefs about rules) is a relic of an outmoded conceptual framework ill-suited to explanation. In what follows, though, I will take for granted that at least sometimes beliefs may legitimately explain.

can contribute to explanations). We might, for example, explain someone being pulled to the side of the road by the police by appeal to the fact that he broke the speed limit and this explanation might go through even though the driver is ignorant of the law. His pulling to the side of the road will be explained partly by his having broken a rule about which he has no belief.

This particular explanation may initially seem plausible only to the extent we appeal to someone else's belief, for instance a police officer's belief that the speed limit was broken. But there are two points to keep in mind. First, the fact that the driver broke a normative rule -- and the fact's explanatory relevance -- is completely independent of the *driver's* recognizing that he was breaking the law. Second, the fact that the driver broke the speed limit might figure in the explanation *even if neither the officer nor the driver believe the speed limit was broken.*

Consider this case: imagine the officer has been ticketing people all day for speeding and has as a result become especially sensitive to those going faster than the limit allows. It is not that she thinks of them as speeders (she usually doesn't, unless and until she clocks them on the radar); it is just that she is disposed to notice them.¹⁵ Suppose that, her awareness of speeders heightened, she notices our wayward driver and the fact that his car fits the description of a car recently stolen. Suppose too that because the officer thinks the car suspect, she takes chase without even thinking about whether the driver was speeding. In this case, the officer will have pulled the driver over because she believed he was driving a stolen car and not because she believed he was speeding (she didn't). Nonetheless, part of the explanation for why the driver was pulled over will be that he was speeding -- had he not been, the officer wouldn't have taken notice of the car.

So rules might play any of the following three roles in explanations: (i) they might be the content of beliefs that in turn explain actions, (ii) they might themselves explain the beliefs (about rules) that explain actions, and (iii) they might explain actions, events, processes, or situations unmediated by beliefs about those rules. Unlike the first role, which leaves actual rules completely out of the explanatory picture, both the second and third roles have the rules themselves doing real explanatory work. The rules, and not merely peoples' beliefs about the rules, are accounting for events in the world.

It might seem as if the three roles are presented in successive order, from that allowing rules the least explanatory involvement to that allowing them the most. But there is no appreciable difference, in that respect, between the second and third roles. Once it is granted that rules can explain behavior when

mediated by beliefs about rules, on the grounds that they explain the beliefs, it is no extra step at all to say they can explain things without the mediation of belief. *For the role a given rule will play in explaining a person's beliefs about that rule will of necessity be one itself unmediated by that person's belief about that rule.* While you can explain my belief concerning some law by appeal to that law, you can't explain my belief by appeal to the belief being explained. Still, it is important to recognize that rules (unmediated by belief) can explain things other than beliefs about those rules, not least of all because in this capacity rules frequently have their subtlest effects.

Although related, the second and third roles do differ -- most significantly in that only when people have beliefs about what a rule requires can they *intentionally* obey or violate the rule. One can, of course, conform to or violate a rule without realizing it; but one can intentionally respect or flaunt a rule only if one is aware of the rule. Unrecognized rules, although they may sometimes help to explain what happens to people and what people do, cannot explain people consciously breaking these rules.

This difference suggests a stronger claim: explanations that give a rule the third role can account only for behavior that conforms to the rule, not behavior that violates it; while explanations that give rules the second role (that is, that invoke rules whose effects are mediated by beliefs about those rules) can explain behavior that violates the rule in question as well as behavior that conforms to the rule. One can of course break the speed limit, say, without having any beliefs about the speed limit, but it may look as if the fact that there is a rule setting the limit won't be relevant to explaining one's violation (even if it is relevant to the explanation of what happens once the rule has been broken). A rule that is neither recognized nor followed by a person may seem beside the point when it comes to explaining her breaking the rule.

But this stronger claim is simply too strong. It is easy to imagine cases where an unrecognized rule might explain its own violation, as when, for instance, a person wishing to stand-out from a crowd takes the necessary steps, only to discover (to her regret) that the crowd was simply obeying a strictly enforced rule. Here, the rule explains the crowd's behavior, and the crowd's behavior (along with the person's desire to stand-out) helps to explain her violation of the rule as well as her consequent woes.¹⁶

To press the point further, even a rule that is unrecognized by everyone (who fall under it) and followed by no one, might explain events in the world if, consistently, those who followed the rule (if ever there were any) would be, for

¹⁵ Such a disposition might, sometimes, reasonably be thought of as a (perhaps subconscious) belief, but not always. We may legitimately suppose, of the case at hand, that the disposition of the officer finds its expression unmediated by her cognitive states (even though she would not have acquired the disposition in the first place had she not had various beliefs).

¹⁶ Something like this apparently underlies the refrain: 'No wonder you got hurt, what you did was stupid!'; the idea being that, even if you didn't know that what you were doing was stupid, its having been so explains your pain.

instance, better off than those who violate it.¹⁷ In such cases, at least part of the explanation of why the people are not better off is that they violate the rule.

More generally, but along the same lines, a rule will be part of a legitimate explanation if the actions the rule enjoins, or forbids, constitute causally significant categories; and it will be an especially informative explanation if the set of relevant actions are otherwise heterogeneous. Explanations of this sort are made most plausible, certainly, when there is some prospect of accounting for the causal feedback loop that renders reliable the effects of conforming to or violating the rule. However, even without a clear picture of how the feedback loop might work, if the effects are sufficiently reliable, and the pattern sufficiently evident, the rule will play a role in reasonable explanations.¹⁸

Explaining by Transconventional Rules

That conventional rules might explain what happens in the world is not, I take it, really all that controversial. In contrast, there are some apparently significant differences between conventional rules (of law, grammar, or games) and trans-conventional rules (of logic, rationality, or morality) that make much less palatable the suggestion that trans-conventional rules might explain.

The first difference has to do with what has become a standard test for explanatory import -- the counterfactual test, which asks "what if the rules had been different?" and counts the rules as explanatorily important only if their being different would make a difference.¹⁹ When we subject conventional rules to the counterfactual test, the test is conceptually unproblematic (since we can make sense of the rules being different than they in fact are). More impressively, it is quite clear that conventional rules can in principle pass the test; often, if the rules had been different, peoples' beliefs (and their resulting behavior) would have been different as well.

¹⁷. This really is just a 'for instance'. The explanatory force of a rule doesn't depend on it having salutary effects when followed. As long as following a rule would have effects consistently different from those violating the rule has, the rule will be part of the explanation of why the effects are (or are not) in evidence. Yet the further away the effects are from being obviously valuable the less normative the rules in question will seem.

¹⁸. The importance of a patterned history, when it comes to explaining by appeal to rules *when the rules are not recognized by anyone involved*, does mean that in these situations the rules cannot provide a single-shot explanation of particular events -- they can explain a single event only by relying on a pattern of events that the rule explains. See Alan Garfinkel's Forms of Explanation, (New Haven: Yale University Press, 1981); Richard Miller's Fact and Method, *op. cit.*; and for a slightly less lenient view, Jon Elster's Ulysses and the Sirens, *op. cit.*

¹⁹. This counterfactual test goes hand-in-hand with the counterfactual analysis of causation. See David Lewis' "Causation," *Journal of Philosophy* (1973), pp. 556-567; and his "Counterfactual Dependence and Time's Arrow," *Nous* (1979), pp. 455-476. See too my "Moral Theory and Explanatory Impotence," *op. cit.*, and Nicholas Sturgeon's "Moral Explanations," in Morality, Reason and Truth, edited by David Copp and David Zimmerman (Totowa: Rowman and Allanheld, 1985), pp. 49-78.

The second difference turns on the fact that both the metaphysical status and actual character of particular conventional rules is fairly unproblematic. People generally agree as to what the rules are and they share a sense of how to investigate and resolve differences of opinion when they arise.

On both counts, things get messy quickly when it comes to trans-conventional rules. First, in many cases it is hard even to make sense of the counterfactual test since it is hard to make sense of the trans-conventional rules being different than we take them to be. This is largely because the rules often don't seem to be contingent in a way that allows us to imagine easily and with any confidence what things would be like if they were in any significant way different. Suppose it were morally permissible for people in our situation to torture young children simply for amusement... what would people believe under the circumstances?²⁰ This seems similar to asking one to suppose that two plus two didn't equal four and then asking what people would believe under the circumstances. One is inclined to say these questions don't have answers because their suppositions are necessarily false, perhaps even inconceivable -- a fact that makes it difficult, to say the least, to apply the counter-factual test.

Second, both the metaphysical status and the actual character of particular trans-conventional rules -- especially those of morality, but also those of rationality -- are very problematic. People frequently disagree as to what the rules are and also as to how one should investigate and resolve the differences of opinion that inevitably arise.

Despite these striking differences, sense can be made of trans-conventional normative rules (if there are any) contributing to our explanations of non-normative events, processes, and situations. To a great extent, the very real differences mentioned above, while they plague any attempt to defend one particular trans-conventional normative theory as over against another, leave largely untouched any particular conception's claim to explanatory import.

Recall that we already have on-hand a couple of plausible explanations that appeal to trans-conventional rules of rationality: the explanations of conventional rules mentioned early-on that made use of decision and game theory. There is room clearly to accept the decision and game theoretic explanations of conventional rules while at the same time rejecting as an inadequate normative theory the maximizing conception of rationality. Conversely, one might accept as right-headed the normative theory while doubting the success of the decision and game theoretic explanations. Yet for those who do accept both the normative view and the explanations (as many do), the decision and game theoretic explanations provide a nice example of normative explanations. The understandable controversy that will surround, say, the maximizing conception of rationality should warn devotees away from

²⁰. Notice that the relevant counterfactual does not ask us to suppose merely that people believe it morally permissible to torture for amusement, but instead to suppose such a belief true.

confident dogmatism; but to the extent they have reasonable (even if not conclusive) grounds for their view, there is no reason for them to refrain from using their normative theory in explanations. And, assuming the explanations are successful, there is no reason for them to refrain from seeing the normative theory they embrace as having explanatory credentials.

When it comes to trans-conventional normative rules, any attempt to develop a normative theory that will hold some promise of contributing to explanations will involve two steps. The first is to get some defensible account of what the relevant rules (of logic, rationality, or morality) are. This is the job of justificatory theory. Only if this first step is taken will the theory have any claim to normative force. Taking the first step usually involves defending one more or less controversial account among many. But it is a step people regularly take. The second step is to show that -- on that preferred understanding of the normative rules -- the fact that some inference, choice, or action, violates or accords with those rules (so understood) explains something outside the inner circle of the normative theory in question -- i.e. explains why the inferences, choices, actions, were made or taken, or explains some other non-normative facts. Only if this second step is taken will the theory have any claim to explanatory force. Neither step is trivial. Yet, if the two steps can be taken successfully, one will be in a position both to justify one's account of the norms and to explain things that happen in the world by appeal to those norms.

It is worth noting that, as long as we are concentrating on trans-conventional rules, taking the second step involves making plausible the view that there is some nonconventional causal feedback loop that is responsive to the difference between what violates the rules and what accords with them. In the absence of some such nonconventional feedback loop the rules will have no role in explaining events independent of their recognition within convention. Significantly, though, when the explanation by rules is backed by a feedback mechanism, the problem involved in applying the counterfactual test to non-contingent rules doesn't arise; for the relevant counterfactual doesn't have to do with what would have happened *had the rule been different*, but with what would have happened *had the behavior being explained been different* in such a way as to make it conform with the rule rather than not (or vice versa).

Not only is it plausible to think defensible normative rules (even trans-conventional ones) might figure as the contents of peoples' beliefs; it is reasonable as well to suppose that, with a suitable conception in hand, one might have reason to think, first, that peoples' beliefs are sometimes responsive to the rules (as when the beliefs in question are one's own), and even, second, that peoples' prospects may vary, other things equal, according to whether they live up to the rules' demands or not. The maximizing conception of rationality is a clear case in point; for it flows naturally from that conception that those who fail to maximize expected utility will, on average, do worse than they otherwise

would have. One doesn't have to look far, though, to discover plausible conceptions of morality that fit the bill as well.²¹

Two, in particular,² are worth mentioning. One is a broadly utilitarian conception, the other broadly contractarian. Both -- despite their deep and dramatic differences -- hold the promise of giving morality not just a justificatory role but an explanatory role.

The utilitarian theory I have in mind begins with a conception of objective human good that is grounded in the satisfaction of informed preferences, and does not simply equate what is objectively good with what people happen to prefer. It then construes moral requirements as reflections of the demand that human good be maximized, without regard to whose good it is that is being advanced. On this view, actions, practices, and institutions, as well as particular moral claims, are seen as morally justified to the extent they contribute to the maximization of human good.

Within this view, so sketched, there is obviously a great deal of room for different specific views of morality, since a wide variety of accounts might be used to fill in more precisely the notion of objective human good. And there is room as well to embrace just the broader consequentialist outlines and, for instance, replace the appeal to objective human good with a purely subjectivist account that identifies human good with the satisfaction of human preferences whatever they happen to be (regardless of whether they are fully informed or reflectively held).²²

The contractarian theory, in contrast, begins not with a conception of objective human good, but with a theory of rationality and, specifically, rational agreement. It then construes moral requirements as the products of rational agreement; the moral rules are those to which rational people would rationally agree, and with which they would then rationally comply, under appropriately specified conditions.

Within the contractarian view, too, there is a great deal of room for different specific views of morality, in this case because a wide variety of accounts might be used to fill in more precisely the notions of rationality and rational agreement. And there is room as well to embrace only the broadest

²¹. Although what follows is limited to a discussion of moral theories and their potential explanatory force, the problems faced by moral theory are exactly paralleled by problems faced by any attempt to defend the explanatory force of a normative system of logic or of practical rationality.

²². This sort of view goes back at least to Bentham's Introduction to the Principles of Morals and Legislation (New York: Hafner, 1948). For variations on this general theory, along with defense of its explanatory value, see Peter Railton's "Moral Realism," *op. cit.*; Richard Boyd's "How To Be A Moral Realist," in G. Sayre-McCord (ed.), Essays on Moral Realism, (Ithaca: Cornell University Press, 1988), pp. 181-228; David Brink's Moral Realism and the Foundations of Ethics, (Cambridge: Cambridge University Press, 1989); and Richard Brandt's A Theory of the Good and the Right, *op. cit.*.

contractarian outlines and, for instance, replace the appeal to appropriately specified, presumably hypothetical, conditions with simply the actual conditions real people happen to find themselves in.²³

Despite the truly fundamental differences, both utilitarianism and contractarianism (of the sorts described²⁴) maintain an over-all structure that makes the theories suitable candidates for use in explanations. For in each case, the demands of morality would be specified in a way that would make plausible the contention that there is a causal feedback mechanism that, by and large and in standard conditions, rewards compliance and penalizes non-compliance with (what each takes to be) the demands of morality.

Thus, for instance, on both the utilitarian and the contractarian accounts one might expect resentment, as well as dissatisfaction with the *status quo*, to grow roughly in proportion to injustice. Within societies (and also perhaps among societies) one could predict also that the social and political pressures would rise and fall in reaction to morally relevant changes in social institutions, consistently applying an impetus for moral improvement, if only with partial success. At the very least, unjustifiable institutions and practices likely suffer instability due to peoples' inability, on reflection, to endorse the norms to which they are subjected. Of course, because the social pressures will vary as well with factors other than the moral credentials of the institutions in place, the sanctions immorality suffers may be buffered, deflected, or simply neutralized. However, that forces will have no easily *predictable* effects in an exceedingly complex environment is no reason to deny their presence or their effect. And, to the extent morality is viewed in either a utilitarian or a contractarian way, there are even some grounds for thinking the effects of immorality are actually quite visible and predictable.

Needless to say, although the two approaches to moral theory each have the resources to make sense of how morality and immorality might be relevant to explaining both large scale and more local events and patterns of interaction, their differences will ensure that their various resources will be deployed in

drastically different and fundamentally incompatible ways. They can't both be right about the nature of morality and so can't both be right about the ways and extent to which morality helps to explain social change. The important point for our purposes, though, is that either sort of moral theory -- as long as it can be defended as a reasonable normative theory -- will hook-up with our experiences, prospects, and fortunes, in such a way as to give morality itself, and not just beliefs about morality, a role in explaining what happens in the world. Either will give sense and substance to the notion that morality operates at least as a gentle breeze through the course of history.

Having spent all this time trying to defend the suggestion that morality might actually contribute to empirical explanations, a confession is in order. Much as I believe morality (at least on some plausible conceptions) does indeed have some explanatory force, I suspect as largely, perhaps even dangerously, misguided the view that moral theory's respectability depends on its having such force. The legitimacy of moral theory rests, I believe, not on its explanatory, but on its justificatory, force. But that is an argument for another time.²⁵

²³ This sort of view has its own obvious roots, most clearly in Hobbes' *Leviathan*, (New York: Penguin Books, 1968). For variations on the general contractarian theme, see David Gauthier's *Morals By Agreement*, *op. cit.*, and his "Why Contractarianism?" in *Rational Choice and Moral Contractarianism*, edited by Peter Vallentyne (Cambridge: Cambridge University Press, 1990); James Buchanan's *Limits of Liberty*, (Chicago: The University of Chicago Press, 1975); and Gilbert Harman's "Justice and Moral Bargaining," *Social Philosophy and Policy* I (1983), pp. 114-131.

²⁴ There is plenty of conceptual space within the paradigms of utilitarianism and contractarianism to allow for other versions that eschew the sort of contact with the world that these theories maintain. The non-natural moral ontology advanced by Moore and Ross, for example, leaves completely mysterious the impact morality is supposed to have on the world. Retreating to the claim that it does so by affecting our moral beliefs via (a special sort of) intuition hardly dissipates the mystery. See G. E. Moore's *Principia Ethica* (Cambridge: Cambridge University Press, 1903) and W. D. Ross' *The Right and the Good* (Oxford: Oxford University Press, 1930)

²⁵ It is an argument I have begun in "Moral Theory and Explanatory Impotence," *op. cit.* This paper was written for, and delivered at, an interdisciplinary conference on "The Logic of Social Change" held at the Murphy Institute of Political Economy, Tulane University, in April 1988. I've since given versions of it at Vanderbilt University, Davidson College, the United States Air Force Academy, West Virginia University, the University of Kansas, and at the University of St. Andrews, Scotland. Thanks are due to all seven audiences. They are due also to Christopher Morris, David Resnik, Michael Resnik, Walter Sinnott-Armstrong and, especially, to David Braybrooke, for detailed and perceptive comments on an earlier draft.