# EVALUATING BAD NORMS

By John Thrasher

Abstract: *Some norms are bad. Norms of revenge, female genital mutilation, honor killings, and other norms strike us as destructive, cruel, and wasteful. The puzzle is why so many people see these norms as authoritative and why these norms often resist change. To answer these questions, we need to look at what "bad" norms are and how we can evaluate them. Here I develop an integrative analysis of norms that aims to avoid parochialism in norm evaluation. After examining and rejecting several evaluative standards, I propose what I call a comparative-functional analysis of norms that is both operationalizable/testable and nonparochial, and that can sort better and worse norms. One conclusion of this approach is that norms are not so much "bad" and "good" as "better" and "worse." This approach should be of interest to theorists and practitioners alike.*

KEY WORDS: bad norms, social norms, harm, social science methodology

## I. Introduction

The ability to understand and follow social norms is arguably the most powerful ability humans possess. It is an ability that allows us to live robustly social lives and benefit from huge networks of cooperation. The norms we live by are diverse and numerous, varying over time and space. Despite this, we tend to be convinced, at any given time and place, that *our* norms are *the* norms—the only proper norms. This feature of norm identification—norm parochialism—is a common and probably necessary part of our psychology that allows us to easily identify, internalize, and police norm related behavior. Norm parochialism is often combined with a tendency to moralize norms, that is, to see them as nonconventional and serious.

Most of the time this combination of parochialism and moralization is innocent enough. We may think that only barbarians use knives and forks in a particular way or that the clothes they wear are scandalous or outrageous. We may also believe that many of our social norms for treating one another, for instance, taboo terms or norms of deference, are backed up by our general egalitarian or cosmopolitan worldview. "Surely any right thinking person would not use *that* word" since it is offensive and hurtful. We may recognize that our social morality is WEIRD (Western, Educated, Industrialized, Rich, and Democratic), without thinking that there is anything wrong with the kind of parochialism involved in western, liberal aversion to racism or sexism.[1]

[1] For a discussion of WEIRD morality, that is, the social morality shared by the educated classes of western democratic society, see Jonathan Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York: Vintage, 2013), chap. 5.

doi:10.1017/S0265052518000055

Parochialism and moralization combine into a more potent stew in the case of so-called "bad norms." These are norms that require and regulate practices that seem destructive or odious. It is surprisingly hard to clearly define how to identify a bad norm, as I will argue below, but some paradigm cases range from the merely inefficient (such as giving gifts at Christmastime), to the mildly dangerous (such as doctors wearing ties), to the horrific (female genital mutilation or so-called "honor" killings).[2] All of these cases involve some inefficient or destructive practice that is, nevertheless, a stable norm.

These kinds of norms are extremely important to understand, both practically and theoretically. Bad norms call out for change, but this is difficult when the members of the norm culture are committed to those norms. We may think that these individuals are in the grips of a set of parochial and moralized norms, but we are also surely in a similar situation. How can we evaluate and advocate for the change of bad norms without ourselves being parochial? Answering this question leads into other questions about the general evaluation of norms.

The variety of norms and social practices around us is considerable, but not infinite. Some norms seem to be more stable and to spread more easily than others. The ultimate end of this project is to develop some way of evaluating this panoply of norms. The challenge is that all of us are already within an existing network of norms. Our norms seem natural to us, other norms seem alien. The important question is whether there is any stable point of view or evaluative standard from which we can judge all social norms, in the same way that we might, for instance, be interested in developing a political conception of human rights that can be universally applied without being merely parochial, or a conception of justice that is stable in the face of considerable diversity.

This project is more difficult than one might initially think. Nevertheless, it is crucial since it is essential to the project of evaluating and, hopefully, changing norms in a more tolerant and humane way. The first step is to be clear about what social norms are and how they work. Then I will look at several promising approaches to evaluating norms, rejecting all of them as flawed in important ways. This leads to questions about how to analyze norms, which will ultimately lead to a method of evaluating norms that can avoid parochialism and point in the direction of effective norm change.

[2] On the inefficiency of Christmas gift giving, see: Joel Waldfogel, "The Deadweight Loss of Christmas," *The American Economic Review* 83, no. 5 (1993): 1328–36; Pedro-Jose Lopez et al., "Bacterial Counts from Hospital Doctors' Ties Are Higher than Those from Shirts," *American Journal of Infection Control* 37, no. 1 (2009): 79–80; John Thrasher and Toby Handfield, "Honor and Violence: An Account of Feuds, Duels, and Honor Killings," *Human Nature* (Forthcoming, 2019).

## II. Norms: The Good and the Bad

There are several different accounts of norms available in the philosophical, economic, and sociological literature. The most influential across these disciplines is the account of norms developed by Cristina Bicchieri and her collaborators.[3] Put simply, there is a social norm for some population if enough people in that population have a conditional preference for following the norm when they expect general conformity with that norm, both empirically and normatively.[4] Under these conditions, norm-following is a Nash equilibrium since no one does better by unilaterally deviating from the norm, either because they would lose from coordination failures or because they would be liable to sanction from others in the group for not complying. The key element is that my preference to conform to the norm is conditional on the belief that others have the same preferences and that others are expecting me to conform. Both of these need to be present in the case of social norms. If others expect me to conform to a norm, but my preference to conform is not dependent on whether others conform to that norm, the norm in question is what Bicchieri calls a "moral" norm. Together with "personal" norms, we can think of nonsocial norms in this sense as not conditional on the normative and empirical expectations of others.

There are several implications of this account of norms. The most important here is that there is a natural distinction between what we might think of as the *positive* social norms of a population and the *possible* norms of that population. The positive norms are those that actually exist in that population, while the possible norms are those that might exist given some change of conditional preferences or expectations. In this way, we can separate the *existence* of a norm from its *evaluation*. We can recognize that a norm requiring racial discrimination exists without thereby endorsing that norm as good. The important point is that it is possible to have genuine norms that are also, in some sense, bad. That is, we can negatively evaluate a norm without thinking that the norm is necessarily defective *qua* norm.

This accords with at least some of our common responses to certain norms. We find culinary norms in some other societies disgusting or bad — specifically, when other societies have norms about what kinds of animals are considered food. In the west, we keep dogs as pets and treat them like furry children. In Korea, China, and many other countries,

[3] See Cristina Bicchieri, *Norms in the Wild* (New York: Cambridge University Press, 2016); Cristina Bicchieri, *The Grammar of Society: The Nature and Dynamics of Social Norms* (New York: Cambridge University Press, 2006).

[4] For a similar explanation of institutions more generally, see: Ken Shepsle, "Rational Choice Institutionalism," in *The Oxford Handbook of Political Institutions*, ed. R. A. W. Rhodes, Sarah A. Binder, and Bert A. Rockman (New York: Oxford University Press: 2006), 23–38; Andrew Schotter, *The Economic Theory of Social Institutions* (New York: Cambridge University Press, 2008).

dog meat is commonly eaten. Contrast this with Australia, where it is illegal in at least one state to eat dog meat and illegal in all other states and territories to sell it. During the Beijing Olympics, the government of the People's Republic of China had dog meat removed from one hundred twelve "official" Olympic restaurants and they were ordered not to serve dog meat during that time.[5] The Chinese elites surmised that visitors would be repelled by the Chinese dog meat norms and this would reflect poorly on the country as a whole.

There are many other examples of seemingly bad norms. But, if we want to give a precise definition of what constitutes a bad norm, we need to go beyond discrete examples and develop a standard that we can use to evaluate norms more generally. This project is really composed of two subsidiary projects, one conceptual and the other evaluative. Conceptually, we need to identify the features any account of bad norms should have in order to fit the common notion or concept of a bad norm. That is, the general account of bad norms developed should intuitively fit with most of our pre-theoretical notions of a bad norm. This idea can be understood intensionally or extensionally. In principle, either understanding is acceptable, but for theoretical simplicity we can think of this idea as core extensional adequacy. The general idea of a bad norm should include the core cases of bad norms like female genital mutilation, honor killing, cannibalism, child sacrifice, and so forth. In addition, the concept of a bad norm should be operationalizable. That is, it should be constructed in such a way that we can test whether such a norm exists and fit it into our most basic social-scientific frameworks like rational choice theory. This aspect is one of the most appealing features of the general account of norms from Bicchieri with which we began this section, and is a benefit of her theory over other competitor theories.[6]

The third desideratum of an account of bad norms is "nonparochialism." This is by far, the hardest feature to develop and it is importantly linked to the evaluation of norms. From any parochial point of view, we can identify and evaluate any number of alien norms as bad. If, however, we are interested in evaluating norms from a perspective that is stable across a variety of parochial points of view, we will need to establish a more robust standard. In addition to being defective as a conceptual standard, a parochial evaluative standard will also be practically defective. One of the important reasons to develop an understanding of bad norms is to change them for the better. If our evaluative standard is merely parochial moralizing, we can have little confidence that our attempts to change bad norms will be anything more than the coercive imposition of our own

[5] "China Bans Dog from Olympic Menu," *BBC News*, accessed November 4, 2016, http://news.bbc.co.uk/2/hi/7501768.stm.

[6] For instance, "the 'Canberra theory' of norms" developed in Geoffrey Brennan, Lina Eriksson, Robert E. Goodin, and Nicholas Southwood, *Explaining Norms* (Oxford: Oxford University Press, 2013).

norms on others. This is a problem insofar as we are concerned about coercion, but also because these changes will be less likely to stick. Nonparochialism does not mean that we need an ultimate normative or evaluative standard in order to evaluate norms. This would require begging too many important questions and would require the use of a controversial theory of morality or value. Without settling any important questions about morality, we should think of nonparochialism as a standard that is accessible from any given parochial standpoint. In the case of norms, the standard requires members of the norm population to be able from their point of view to see the norm as bad. Of course, they may not regard it as bad now, but they would not need to adopt a foreign point of view in order to see the norm as bad.

As I will argue in the next two sections, to fully make sense of these features we will need to look more at the details of how to explain and evaluate norms. Before filling in the details, though, my claim is that an adequate account of bad norms will have three important features. Conceptually, any account of bad norms should be able to capture the intuitive paradigm cases of bad norms. Theoretically, a particular account should be useful in identifying actual norms within a usable social-scientific framework. The account should also be practically accessible from a nonparochial point of view. These features are identified in Table 1.

The theoretical desideratum of operationalizability is realized by adopting Bicchieri's general approach to norms. She has shown that it is operationalizable in empirical and experimental contexts. Without worrying too much about the meaning of the concept of bad norms, we can at least focus on the core cases and think of the idea of bad norms as being defined extensionally to be those cases. Nonparochialism is the hardest condition to meet, although as we saw above it is possible to weaken the condition slightly by making it more of an accessibility condition. If members of a norm group are capable of seeing that the norm is bad, then it is nonparochially bad.

In this section, I have tried to provide a thin but nevertheless useful account of what a bad norm is. We still don't have an account of what makes the norms bad — that is, how to evaluate badness — but I have given some conditions that any account of bad norms will need to meet. To fill in those conditions, however, we will need to look more closely at how we explain and evaluate norms more generally.

Table 1. *Features of Bad Norms*

| Conceptual | Theoretical | Practical |
|---|---|---|
| Extensional Adequacy | Operationalizable | Nonparochial |

## III. Evaluating Norms

In this section, I look at several possibilities for evaluating norms, rejecting each one, either because of internal problems or because it fails to meet the nonparochialism condition. In the next section, I look more closely at the explanatory structure of bad norms. This will point in the direction of a different approach to evaluating norms that I will assess in Section V.

### *A. Harm*

The first and, perhaps, most straightforward way to evaluate bad norms is on the basis of harm. Surely what makes norms like female genital mutilation bad is that they are harmful. This also seems to be true in other cases like public defecation, honor killing, and norms of revenge.[7] A plausible evaluative standard for bad norms is that bad norms are harmful norms. Bad norms are the norms that either require or are likely to cause harm. We can state this definition of a bad norm as:

> **Bad Norms as Harmful Norms:** A norm *R* is bad if it requires or is likely to cause harm to those in population *P*.

There are several problems with this seemingly attractive formulation. First, it is not clear whether the harm must be likely to accrue only to some members of *P*, or to all the members of *P.* For instance, in cases of female genital mutilation, only women are affected by the norm directly and even when the norm is in place, it is rare for all of the women in the group to be affected.[8] In what sense, then, is the norm harmful to the men? Similarly, a norm of public defecation near a riverbank seems to meet the intuitive extensional conditions to be a bad norm, but it might be that, in a certain case, those who defecate near the river only drink water that has come from upstream and do not end up getting sick since their water is not contaminated. They are, however, contaminating the water downstream from them and are likely to get anyone who is collecting water downstream sick. The harm in this case is imposed on those outside of the population *P.* In these cases, it looks like the formulation of bad norms as norms harmful to those in *P* is not a necessary condition of something being a bad norm.

More importantly, though, harm cannot function as an evaluative standard for bad norms because harm itself is a norm-mediated evaluative

[7] Bicchieri, *Norms in the Wild*; Thrasher and Handfield, "Honor and Violence"; Christopher Boehm, *Blood Revenge: The Enactment and Management of Conflict in Montenegro and Other Tribal Societies* (Philadelphia: University of Pennsylvania Press, 1986); Jon Elster, "Norms of Revenge," *Ethics* 100, no. 4 (1990): 862–85.

[8] In Charles Efferson, Sonja Vogt, Amy Elhadi, Hilel El Fadil Ahmed, and Ernst Fehr, "Female Genital Cutting Is Not a Social Coordination Norm," *Science* 349, no. 6255 (2015): 1446–47.

concept—what counts as a harm will largely depend on the norms of the particular context. To take some trivial but nevertheless illuminating examples, consider male versus female genital mutilation. In the former case, we call it "circumcision" and many western countries do it as a matter of course when children are born and there is little opposition to the practice. The female version, in all of its different forms, has substantially more negative effects and the two kinds of "circumcision" are only similar in that they both involve cutting of the genitals. We might think that the difference is that the male version is considered a negligible harm, while the female version is a more serious harm. In terms of pain and related health complications, this distinction is warranted. The problem, however, is that the male version, where it is the norm, is *not* considered a negligible harm; rather it is typically considered *harmless*. Harming is different from hurting. Something that hurts may or may not be a harm, and harms may or may not hurt. Harming is a kind of wronging and it, consequently, has a deontic aspect. Admittedly, this is not the only way to think of harm, but it is a natural one in the moral context.

Another interpretation is that the male circumcision case is really a case of *justified* harm. Circumcision is a harm, but that harm is justified on the basis of religious or health considerations. Similarly, surgery is harmful, but it is often justified because of the benefits. This example, however, precisely shows why it makes little sense to think of harms in this way. Do we really think that the surgeon is harming the patient by cutting him up? Even if the patient dies on the operating table, it seems odd to claim that he or she was harmed by the surgeon unless, and this is the crucial point, the surgeon acted improperly or negligently, which is to say unless the surgeon violated the relevant norms. Similarly, my suspicion is that Jews do not believe that mohels are harming their boys by circumcising them—quite the opposite.

The role that consent plays in harm highlights the fact that harm is a normatively laden concept. Punching or being punched by someone is generally considered harmful behavior, but if this punching is done within a boxing match or a hockey game, it is not considered harmful (as long as it is within the norms of that game). This is not because we expect fewer negative effects to occur from boxing than from some other type of punching; boxing is not harmful since both parties have consented to engage in the practice in certain norm-governed ways. Consent can play this important role, because harm is a specification of the normative properties and deontic restrictions between individuals.

Rights violations of various kinds can be harmful even if there is no loss. For instance, imagine Tom takes Ralph's car while Ralph is at work and runs an errand with it, returning it to Tom before Ralph gets back from work. Tom didn't have Ralph's permission and Ralph and Tom are not very good friends. In order to sweeten the situation, Tom leaves a twenty dollar bill on Ralph's dash. It seems perfectly reasonable to say about this

case both that Ralph was harmed by Tom and that Ralph was made better off. If this case isn't compelling, imagine some other, more important, trespass that is not consented to, but is also somehow compensated. Someone in a coma, for instance, can be harmed even if he or she never finds out about it and is not made worse-off overall.

Harm relies on various other norms for its specification in particular contexts. This is partially what makes harm such an intuitive and useful concept to use within a particular society as an evaluative standard, but less effective when it goes beyond a given normative framework. Using harm as the standard will not only beg the evaluative question, it will not meet the nonparochialism condition since any harm standard will have some sort of parochial norm system associated with it. It might be possible that there is a standard of harm that is independent of any other parochial normative point of view, but there is no evidence of one as of yet.

### *B. Efficiency*

A thinner evaluative standard that does not require reference to additional norms is an efficiency standard. In this context, efficiency is the economic concept of efficiency as Pareto efficiency. For norms we can think of this standard in the following way:

> **Bad Norms as Inefficient Norms:** A norm $R_1$ is a bad norm if there is some other norm $R_2$ that no member of $P$ would rank as worse than $R_1$ and at least one member of $P$ would rank as better than $R_1$.

There is an obvious advantage to this standard over the harm standard in terms of nonparochialism since the evaluation is made from the point of view of each member of *P.* This advantage, however, comes at a cost for operationalizability.

There are two significant problems that make an efficiency standard difficult in terms of operationalizability. The first is identifying the set of feasible alternatives to make efficiency comparisons. Bad norms on this view are basically norms that are off the Pareto frontier; but we need to have a good idea of where that frontier actually is in order to make determinate evaluations. If all that is required is to show that some alternative norm is preferable to the current norm, this is a trivially easy task to accomplish. Any norm can be specified in such a way as to make it a Pareto improvement over the current norm if it is off the Pareto frontier. The problem is that there will be many, perhaps innumerable, other norms that would also be Pareto improvements, but that might benefit certain persons more than others. In those cases, distributive concerns will arise and we will need a standard other than the Pareto standard to solve these problems. In any case, without precisely specifying the set of feasible alternatives, it will be impossible to define the Pareto frontier and to give a thorough evaluation of the bad norms.

One solution to this will be to make the efficiency standard comparative so that we are not identifying bad norms, but merely identifying better and worse norms. So if members of $P$ would choose norm $R_1$ over $R_2$ on the basis of their preferences, then $R_1$ is better than $R_2$ according to efficiency. This has certain theoretical benefits, but it does even worse on the standard of operationalizability. The reason is that this standard is path dependent and does not translate to the previous efficiency approach. The possible Pareto improvements depend entirely on what the current norm is and what options we have. There are some options on the Pareto frontier that it will be impossible to realize given only Pareto permissible moves. This means that we need to know not only the options but also the paths to realize them. The fact that a norm is Pareto optimal does not tell us anything about whether there is a Pareto permissible path that we can take to realize that option.

Efficiency as a general standard of evaluation, then, is either nonoperationalizable or incoherent as a standard that can meet the conceptual adequacy and nonparochialism requirements. The point is that for efficiency to work, it needs to be highly specific in a well-defined and limited counterfactual space. This is not likely in the case of norm evaluation. Efficiency considerations are also typically path dependent in a number of ways, and this makes them extremely complicated in terms of operationalizability and often parochial as well. Despite this, as I will argue in Section IV, a modified version of the efficiency standard, when combined with other elements can form the basis of a standard of norm evaluation.

### *C. Welfare*

Once we see the problems with the normatively thin efficiency standard, we might be tempted to move toward a more substantive welfare standard. By invoking an objective standard of welfare, we could potentially evaluate whether the norms in question are welfare enhancing or not and define bad norms as those that are detrimental to welfare.

> **Bad Norms as Welfare Reducing Norms:** A norm $R_1$ is a bad norm if the net result of $P$ of $R_1$ is less welfare than would result if $R_1$ were not in place.

Formulated in this way the standard is both counterfactual and comparative. We have to evaluate whether the member of $P$ would be better off with or without $R_1$ and the counterfactual world may include some other norm or no norm. So, in effect, we are comparing $R_1$ to some other feasible norm $R^*$ and no norm $R_0$ along a welfare metric. If $R_1 > (R^* \vee R_0)$ then we can say that $R_1$ is not a bad norm.

There are several problems with this formulation. The first is that, like the efficiency standard, we need a discrete and reasonably small set of

*R** in order to be confident about whether the norm in question is bad. It will be trivially easy to show that any norm is bad if merely any other norm can be shown to be better in some welfare terms. Second, since the evaluation is counterfactual we need to be able to keep all the non-norm features fixed when we are doing the counterfactual analysis. Often this is very difficult to do since the norms are embedded in complex cultural and social networks that make changing the norm and only the norm difficult or impossible. This is a practical as well as a theoretical problem that impacts the operationalizability of this approach. A similar problem arises with Randomized Control Trials (RCT) and the Instrumental Variable (IV) approach in development economics. As Angus Deaton has argued, both of these approaches have a similar problem when results are applied across time and between societies.[9] We should be skeptical of our ability to make very good evaluations or inferences on the basis of this standard in an operationalizable way.

These objections all assume that the welfare metric is itself something that we have confidence in, but there is little reason to think that we can find a good welfare metric to play that role effectively. Philosophical theories of welfare are as numerous as welfare theorists, and most are too abstract and formally thin to be very helpful in the actual evaluation of real norms. They also tend to be parochial. Any general welfare metric will also have the problem of requiring some interpersonal comparisons of utility, which there is a long tradition of seeing as illegitimate.[10]

Despite these problems, we might still think that there is some welfare proxy that we could use, and following development economics there seem to be several plausible candidates—GDP, for instance, or the more disaggregated purchasing power parity (PPP) measure to evaluate how well-off different societies are in comparison to one another. There are good data on PPP measures, and recent work has led to significantly improved versions of them.[11] The problem with using these measures with norm evaluation is that they are too coarse-grained typically to allow for comparative norm evaluation. There are many factors that contribute to GDP and PPP in complex ways—so many that development economists have no clear theory about how they interact with culture, history, institutions,

[9] See: Angus Deaton, "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature* 48, no. 2 (2010): 424–55; Angus Deaton, *The Great Escape: Health, Wealth, and the Origins of Inequality* (Princeton, NJ: Princeton University Press, 2013).

[10] Lionel Robbins, "Interpersonal Comparisons of Utility: A Comment," *The Economic Journal* 48, no. 192 (1938): 635; Kenneth J. Arrow, *Social Choice and Individual Values*, rev. ed., (New Haven, CT: Yale University Press, 1963).

[11] Angus Deaton, "Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll," *The Journal of Economic Perspectives* 22, no. 2 (2008): 53–72; Angus Deaton and Olivier Dupriez, "Purchasing Power Parity Exchange Rates for the Global Poor," *American Economic Journal: Applied Economics* 3, no. 2 (2011): 137–66.

geography, and norms. We might be able to use these measures as crude proxies for societies that are generally not well-off, but we won't be able to know if they aren't well-off because of or despite a certain norm. Another possibility is to use subjective measures of well-being or happiness, but this has many of the same problems of being too course-grained and information intensive, while also replicating some of the problems from efficiency standards.

These problems should lead us to look for another solution, an approach that can be both operationalizable and nonparochial and can capture the intuitive difference between good and bad norms. To understand what is needed, however, we must look at how we explain bad norm following. Doing so will lead to a method of analysis that we can use to develop an evaluative standard that avoids some of the problems discussed here.

## IV. Explaining Bad Norms

If we have identified some norm that we think fits the bill as a bad norm we also need to explain how the members of the norm group could both see it as a norm and potentially as a bad norm. That is, if the members of the group already mostly don't agree with the norm or follow it, it won't meet the basic existence conditions of a norm. Instead, we need to show that even though the norm may be bad, members have some reason to follow it as a norm. For instance, in some cultures norms of revenge and feuds are common.[12] These are core cases of bad norms in the conceptual sense from above. At least a critical mass of people within these societies, however, see these norms as *norms*, that is, as creating legitimate normative and empirical expectations. We need a method of analysis that recognizes this datum and incorporates it into the analytical and explanatory framework.

One might think that this approach is merely a descriptive or positive analysis of norms that ignores the properly normative aspects of evaluation. If this were true, it would be fatal to the project presented here since the ultimate goal is to *evaluate* norms, not merely to describe them. Of course, many would dispute a strong fact/value or descriptive/normative divide, but is not my goal to make any important claims on that issue here. Still, it is worth noting that this distinction is especially flexible in the case of norms. Norms are, not surprisingly, the basic material of normativity and when we are analyzing norms our task is both descriptive/positive and normative. The important point for the investigation here is that when we are dealing with genuine norms, the members of the norm group, at least, see themselves as having reason to follow these norms. Insofar as the theorist is concerned with practice and operationalizability—as I am—this

[12] Boehm, *Blood Revenge*.

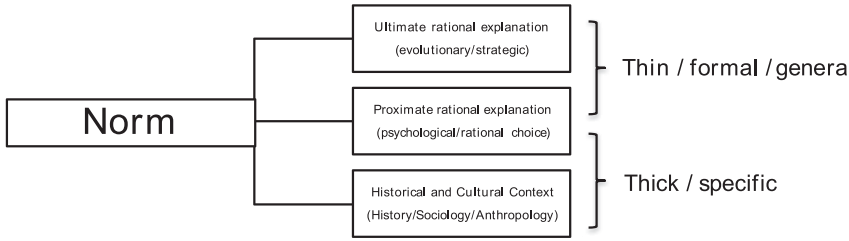
cannot be ignored and the normativity of the norms for the norm population must be taken as a given. We can do this by adopting something like a Quinean or Davidsonian *principle of charity* when interpreting bad norms. That is, we should interpret the norm so as to construe the followers of that norm as rational. The assumption here is that all normal human agents are rational in the minimal sense of behaving in coherent, nonrandom ways. Our job is to rationally reconstruct what could be the coherent reasons why these individuals are following this particular norm and what, in turn, the norm is.

Although it is a basic assumption of most social science, this principle has some interesting properties in the present context. It means that all good explanations or analysis of norms will have a complex structure. Since it is always possible to show that someone had a reason to do something, in this case to endorse or follow a norm, on the basis of some set of beliefs about the world, we need a nontrivial way of fixing and evaluating those beliefs. To do that effectively, we need to evaluate norms on three different, though related levels. The first is the *proximate* level and concerns why an individual has reason to endorse or follow a norm. These explanations will be psychological or have rational choice explanations. The types of explanation in most of the economics literature have this form. We can still ask, however, why *this* norm and not another. And why will individuals have evolved or developed preferences to follow these norms. To answer these questions, we will need to move to the second level of evaluation, which targets the norm itself and specifically its stability properties. These *ultimate* explanations look at how this particular norm could have evolved and stabilized. Once we have these two parts of the explanation, we also need to understand at a third level of evaluation the details of how the norm in question works in a particular historical or cultural context. To do this we will need to look at ethnographic and historical data. This will help us know if the previous two levels of our analysis are well founded. These three levels are described in Table 2.

So, following Mackie, the ultimate explanation of female genital mutilation in Africa is that it is a norm equilibrium that creates confidence in paternity and prevents women from being unchaste.[13] A sociobiological story can be told along these lines and Mackie and others have shown how this could be a stable equilibrium using plausible assumptions. Using historical and ethnographic data, we can test this hypothesis against the history of the peoples who practice female genital mutilation. We can then hypothesize about the proximate rational explanation in terms of the beliefs and preferences the men and women in the norm group must have in order to maintain the practice. This can be further checked with interviews and surveys.

[13] Gerry Mackie, "Ending Footbinding and Infibulation: A Convention Account," *American Sociological Review* 61, no. 6 (1996): 999–1017.

TABLE 2. *Integrative Norm Analysis*




At the proximate level there are several types of explanations for why particular individuals follow a particular norm. The most deflationary answer is that they prefer to follow the norm than to do otherwise. This explanation is a trivial one in one sense, but nontrivial in another. There is the question of whether individuals follow norms because they are norms; that is, do they have a preference to follow a norm or is the norm an emergent property of their preferences taken as a whole? The second is the case of classic Lewis style conventions or pure coordination problems.[14] In a Lewis convention, a practice is a convention not because every individual recognizes that he *should* follow the convention, but merely because knowing the practice allows him to coordinate effectively given that everyone else does the same. Insofar as there is a normative element in conventions, it is a tremendously weak form of rational normativity. The interesting thing, though, is that it is very difficult to find uncontroversial cases of conventions of this form. The most used example, perhaps, is driving on the left or the right, but I suspect most of us would imbue this with a stronger normative force than the conventional explanation should allow. If we were to see someone driving on the wrong side of the road we wouldn't tend to think "look at that fool, he would coordinate better if he were only on the other side" but rather, "look at that dangerous lunatic who is on the *wrong* side of the road, putting us all at risk!" As someone who has recently transitioned from driving on the right to driving on the left, I can testify that the transition was not only difficult, but actually felt *wrong*. It seems to be very easy to load our conventional expectations with normative heft generating genuine social norms. Paul Rozin's description of the process of moralization may shed some light on this phenomenon.[15] We may also want to look to Haidt's social intuitionist

[14] David Lewis, *Convention: A Philosophical Study* (Cambridge, MA: Harvard University Press, 1969).

[15] Paul Rozin, "The Process of Moralization," *Psychological Science* 10, no. 3 (1999): 218–21.

model of moral judgment.[16] According to Haidt, our judgments are preloaded and are heavily affect laden. When we give reasons for our judgments we do so to convince others, not to justify those conclusions to ourselves. Since normative expectations and evaluations are some of the most powerful weapons in our persuasive and justificatory arsenal, it is perhaps no surprise that we are quick to deploy them. What is more surprising is that we should be so susceptible to their charms.

Sometimes this normative projection can cause problems, however. One example is in the case of pluralistic ignorance. This kind of situation creates what Bicchieri calls a "collective illusion" about a norm and, as she notes, norms based on collective illusions can be "fragile."[17] Once the illusion is undermined, there can be a cascade away from the norm. One strong hypothesis about why people follow norms that seem obviously "bad" like FGM is that there is widespread pluralistic ignorance. As Mackie is careful to point out, almost all of the people who practice FGM claim to love and want what is best for their daughters.[18] They believe everyone else in their group wants the same. The problem is that they have false beliefs about how much other people are committed to the norm. In reality, so this explanation goes, most people are opposed to the norm and only follow it because they wrongly believe that others are committed to it. Although there is some reason to think that this original explanation was too simple, Bicchieri takes the idea further and introduces a variable *K* that acts as a weight on how much the norm in question factors into an individual's utility function.[19]

The pluralistic ignorance explanation is charitable since it assumes that individuals in a norm group are rational and it seeks to formally characterize their utility functions. It does, however, assume that those individuals are under an illusion and that the norm will evaporate once this illusion is punctured. Another explanation is that, given the alternative available, the norm in question is the best available or, at least, that there is no better norm available. If this is a possibility in a large number of bad norms cases, as I will argue that it is, then pluralistic ignorance is a special not general proximate explanation. The implication is troubling since changing bad norms will be more difficult than undermining an existing illusion. This will lead us away from proximate analysis on its own and toward linking proximate and ultimate analysis with an evaluative standard. I turn to that in the next section.

[16] Jonathan Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108, no. 4 (2001): 814–34; Haidt, *The Righteous Mind*.

[17] Bicchieri, *The Grammar of Society*, 181; John Thrasher and Kevin Vallier, "The Fragility of Consensus," *European Journal of Philosophy* 23, no. 4 (2015): 933–54.

[18] Mackie, "Ending Footbinding and Infibulation."

[19] Sonja Vogt, Nadia Ahmed Mohmmed Zaid, Hilal El Fadil Ahmed, Ernst Fehr, and Charles Efferson, "Changing Cultural Attitudes towards Female Genital Cutting," *Nature* 538 (2016): 506–9; Bicchieri, *The Grammar of Society*, chap. 6.

## V. Comparative-Functional Analysis

The main problem with the welfare approaches is that they attempt to present a global standard to evaluate norms. As we saw, these approaches had several serious problems, as do all other methods of attempting to provide a global evaluative standard for norms. In addition, we saw in the last section that understanding norms requires different kinds of analysis at several different levels. Most analysis is concerned only with proximate or historical/cultural evaluation. In this section, I will argue that we need to incorporate more of the ultimate analysis in order to evaluate norms and I show how this can be done in the case of honor norms.

To move beyond the evaluative standards we saw in Section III, we need to expand the information base we use to evaluate norms. The efficiency and welfare approaches only look at outcomes and ignore the process by which norms generate these outcomes. This may seem harmless since we are ultimately concerned about the welfare effects of these norms, but it makes our evaluation of these norms more difficult as well. In this section, I will argue that we need to look at the functional role that norms play in particular contexts to be able to coherently evaluate them. Once we see what roles these norms play, we can attempt to evaluate whether there are alternative norms that can also play that role as well or better than the norm in question. If so, we can then evaluate what the benefits and costs of shifting to a new norm will be along various dimensions. In some cases, the new norm will dominate the old norm on all the relevant dimensions and in those cases this alternative norm will be preferable to everyone. This is not the typical case, but whatever the different costs and benefits, this form of evaluation will make it clear to those participating in the norm how expensive the norm is in terms of the opportunity costs of switching to a new norm.

Put simply, this comparative-functional approach to evaluating norms solves some of the problems that we have seen with other ways of evaluating norms. First, it avoids the Pareto problem by not attempting to make a global evaluation and thereby avoiding the problem of fixing the context of comparison. The Pareto standard requires a fixed standard for evaluation in order to determine whether something is a Pareto improvement or not. This parameter is free in a global standard, making Pareto evaluations of norms either trivial or impossible. On the local functional approach, however, we can fix the Pareto baseline at the status quo and evaluate the norm in question with respect to a fixed neighborhood of other possible norms. So, rather than evaluating a given norm against all other possible norms, we evaluate a specific norm against a set of feasible alternatives. What determines the feasible alternatives is not always clear, and it will be necessary to look at the functional role of the norm in question to help populate that set.

We should not confuse the idea of a "functional role" with traditional "functionalist" social theory. Functionalism as a general approach to social

theory has rightly been discredited.[20] Instead, the "functional" aspect of comparative-functional analysis should be understood as a kind of as-if analysis that mirrors, at the ultimate level of analysis, the assumption of charity at the proximate level of analysis.[21] That is, we assume that if a norm exists, it represents a stable equilibrium of social interaction. This alone, however, tells us very little. We need to know more about the kind of interaction for which this norm represents an equilibrium in order to understand the norm. As Bicchieri argues, norms often represent equilibria that are only possible once they have been implemented and, thereby, have changed the initial base game. To identify the social "functional role" of a norm, then, is to identify what game this norm acts as an equilibrium for. Only once we know that can we effectively compare the norm in question to other norms that might also be equilibria to that game or to think of ways of changing the game with additional or different norms.[22] Obviously, then, comparative-functional analysis is not a rejection of methodological individualism. Indeed, the integrative approach explained above and comparative-functional analysis only make sense from an individualist point of view. Any "functional role," since it is an equilibrium in a game, must be ultimately reducible to the proximate preferences, beliefs, and values of the individuals in the norm group. Still, comparative-functional analysis goes beyond traditional rational choice theory in unifying the ultimate level of analysis ("as-if" functional role) with the proximate level (the rational decision making of individual choosers). Doing so allows us to determine the set of feasible norm alternatives.

Once this set is populated, we can compare a specific norm $R_1$ against members of $R^*$ and ask whether the move from $R_1$ to each element of $R^*$ would be a Pareto improvement. The members of $R^*$ that would be judged a Pareto improvement for the set $M$, the maximal set of $R^*$ where $M \subseteq R^*$ and $M(R^*, \succcurlyeq) = [x \mid x \in R^* \,\&\, \textit{for no}\; y \in R^* : y \succ x]$.[23] That is, the maximal set is composed of all the elements where there is no element that strictly dominates another. Put simply, $M$ defines the Pareto frontier of feasible, efficient alternatives. The alternatives in the maximal set, by definition, do not Pareto dominate one another and so, from the point of view of Pareto analysis, this is the end of the story. But there is clearly more we can say about each alternative. It might be that some alternatives are comparatively better for some members of the population or that some

[20] See Jon Elster, "Marxism, Functionalism, and Game Theory: A Case for Methodological Individualism," in *Theory and Society*, ed. Derek Matravers and Jonathan E. Pike (New York: Routledge, 2003), 453.

[21] The classic case for this kind of analysis in economic theory is made in Milton Friedman, "The Methodology of Positive Economics," in *Essays in Positive Economics* (Chicago: University of Chicago Press, 1953), 3–43.

[22] I thank the other authors and editors from this issue for pushing me to make this point more clearly, especially Gerry Mackie.

[23] Amartya Sen, "Maximization and the Act of Choice," *Econometrica* 65, no. 4 (1997): 745–79.

alternatives seem like particularly good or bad matches for a given population. Comparing those alternatives to one another and to the status quo norm, we can generate the relative price of the alternatives in terms of the opportunity cost of moving to one norm rather than another. Different members of the population will view this price differently. Some will see the status quo as very expensive while others will see particular alternatives as expensive.

There is an additional question about whether each norm has one and only one functional role or, to put it another way, whether there is a principled way of individuating norms and functions.[24] In one sense, there will always be a certain indeterminacy in the individuation of norms in the same way that there is a fundamental problem of individuating states and actions in decision theory. Or rather, the problem is in providing a unique individuation of norms and, for that matter, functions. To ask whether there is only one function for each norm is really to ask two questions then: First, can norms and functions be uniquely individuated? And second, is there a function that relates each function to each norm? I am not confident that either of these questions can be answered in a nontrivial way that is externally valid. Both individuation and the relationship between norms and functions proceed from the goals of the researcher. This is to say that the practice of investigation will fix the answer to these questions as work continues on these problems. By following the integrative approach, feedback between the local level of norm participants and models will contribute to making the individuation and relationship between norms and functions more and more externally valid over time.

Following the tradition in social choice and bargaining theory, we might see the goal at this point as developing a mechanism with certain properties (strategy-proof, incentive compatibility, and so on) for negotiating these prices and determining a unique solution to which norm should be chosen. This would be a mistake in the case of norms, however. Norms do not arise from a process of collective choice and it would be misleading to think of the process of norm evaluation and change as similar to collective choice rather than social evolution.[25]

This approach has the potential to avoid parochialism because it requires us to look at the norm both from the ultimate point of view and the point of view of those within the norm group. Evaluating a norm from either point of view separately would miss something important, either the function of the norm or the reasons individuals have for following it. Evaluating feasible alternatives and weighing their relative costs also requires us to take the point of view of those within the norm group. All of this should act

[24] I thank Jerry Gaus for raising the importance of this question.

[25] See Gerald Gaus and John Thrasher, "Social Evolution," in *The Routledge Companion to Social and Political Philosophy*, ed. Gerald Gaus and Fred D'Agostino (New York: Routledge, 2012), 643–54.

as a check on parochialism. Further, since this approach basically adopts a Bicchieri norm analysis at the proximate level it should be operationalizable and testable in the same ways. Conceptually, we use our already existing notions of bad norms to identify potential candidates to test, so it should meet that desideratum as well. It looks like the comparative-functional account can meet all of the desiderata set out in Section I.

Used as a method for analyzing and evaluating norms, the comparative-functional approach can be understood as a high-level algorithm or as a series of questions:

1. What is the norm?
2. What social functional role does this norm play?
3. What is the historical/institutional/cultural context of the norm?
4. What are the reasons that individuals comply with and expect others to comply with the norm?
5. Are there other feasible norms that:
   a. would perform the same or similar functional role?
   b. are compatible with the context?
   c. could be supported by similar reasons?
6. What are the relative costs and benefits of moving to one of these norms for those in the norm group?

Answering all of these questions will be difficult and require attacking the problem from all three levels of analysis described in the last section.

To see how this works in detail, it is worth looking at a concrete example of this comparative functional analysis in action. My colleague Toby Handfield and I have developed a comparative-functional analysis of violent honor norms, what we call *honor-based violence*.[26] Feuds, vendettas, and duels undertaken to defend one's honor or to repair some slight to honor are classic examples. Another form of honor-based violence is the honor killings of women who have violated or are believed to have violated community sexual norms. We argue that in societies with weak governance institutions, norms of honor-based violence can help to solve two types recurrent problems: deterrence and assurance. We divide the honor norms dealing with each problem into what we call *revenge* and *purification* honor norms.

Revenge honor norms serve to address one of the oldest problems of politics: how to emerge from a Hobbesian state of nature, where "there is no place for Industry; because the fruit thereof is uncertain," to a society in which property rights are sufficiently secure that economic activity can thrive.[27] This problem—the *deterrence problem*—is the problem of establishing

[26] The full account can be found in Thrasher and Handfield, "Honor and Violence."

[27] Thomas Hobbes, *Leviathan* ed. Noel Malcolm, Clarendon edition of the Works of Thomas Hobbes (Oxford: Oxford University Press, 2012), chap. 13.

a credible threat that violations of one's self or property will be met with violence in virtually all circumstances, regardless of the strength of the attacker. Or, more precisely, it is the problem of ensuring that threats are not evaluated by weighing the costs and benefits of response at a particular time, on a case-by-case basis. Credible deterrence is a combination of two features. First, one must be able to signal strength. Second, one must be able to signal that one (or the collective of which one is a part) can be counted on to carry out retaliation, even when this is not strictly rational. Attacks or trespasses will be resisted even when the costs of doing so will outweigh the benefits.

The second type of honor norm, which we call *purification* honor norms, relate to assurance. The practice of "honor killings" and similar phenomena emerge from this category. Honor killings occur when, typically, male members in a family kill one of their female relatives because she has violated norms of sexual impropriety. In their eyes, she has "dishonored" the family with her behavior and they believe that the only way the resulting debt of honor can be paid is with her life. For instance, in October of 2009, Faleh Hassan Almaleki killed his daughter Noor Almaleki by running her down with his Jeep in a Phoenix, Arizona parking lot. Noor, who was twenty years old, had defied her father by walking away from an arranged marriage with a cousin in Iraq and by living with her boyfriend against her father's wishes. She had, according to her father, become "too westernized."[28] In that Phoenix parking lot, Faleh revved his engine and accelerated directly into his daughter, dragging her for twenty feet behind the Jeep.

Ten years earlier, in 1999, a nineteen-year-old Kurdish Swede named Pela Atroshi was executed at close range by a gunshot to her head.[29] The killer was one of her uncles who, along with a group of his other male relatives, and motivated by the belief that Pela failed to adhere to traditional Kurdish principles of sexual morality for women, plotted and carried out the murder of his niece. Two of these men, the patriarch Abdulmajid Atroshi and one of the accomplices, were living in Australia at the time. They lured Pela back to Kurdish lands in Northern Iraq where they assumed—correctly—that the honor killing would be punished lightly. Three of the murderers were convicted. They were sentenced with a one-year, suspended sentence. The court cited the "defendants' honourable motivation" as a reason to excuse their offense. In societies that practice honor killing, there is a strong preoccupation with a woman's sexual behavior. Women are frequently confined, veiled, and deprived of opportunities to work

[28] Tim Gaynor, "Iraqi Guilty of Murder in Daughter's Honor Killing," *Reuters*, February 22, 2011, http://www.reuters.com/article/2011/02/22/us-arizona-iraqi-idUSTRE71L8IT20110222.

[29] Sian Powell, "Australian Links in Honour Killing of Pela Atroshi," *The Australian*, accessed November 8, 2016, http://www.news.com.au/national/australian-links-to-brutal-honour-killing/story-e6frfkp9-1111116166086.

or study. A superior equilibrium is available, in which these costly practices are not regarded as prerequisites for marriage; but because any family that deviates from the conventions is likely to suffer immediate costs in the marriage market, it is difficult to shift. Another example of norms that seem to have the same function concerns the norms that govern violent gangs in American prisons.[30] In prisons, there is no recourse to external governance mechanisms to deter violence and the prisoners must organize themselves into protective associations to provide that function.

In the case of honor killing, the norm is that the male family members in a culture with this norm must kill a female family member who has been (or is thought to have been) unchaste. The social function the norm performs, we argue, is to guarantee purity and assurance between families in heavily asymmetric and long time-frame exchanges. The historical/institutional/cultural context of the norm varies from society to society, but there are important similarities in family structure and unreliability of paternity. Individuals comply with the norm because they see it as reflecting on the "honor" or value of their family and see the sister or daughter as undermining the reliability or trustworthiness of the family.[31]

Are there other feasible norms in this case? That question is difficult to answer. In one sense the answer is obviously "yes." Marriages occur in parts of India, Pakistan, and other parts of the world without resorting to honor killing, so we know that there is a feasible alternative because there is an actual alternative.[32] The relevant question, though, is whether the people in the norm group would see this option as feasible. Presumably, numerous members of the norm group know that this is not how women are treated in many other parts of the world and yet still engage in this practice. We hypothesize that stronger, civil marriage law, a more open marriage market, and reliable paternity information could help make alternative norms seem less costly. More work still needs to be done on this issue, however, since there are a number of remaining unknowns about the properties of this signaling equilibrium, even if our analysis is correct.

## VI. Conclusion

One implication of comparative-functional analysis of norms is that we should probably be less confident in our intuition that some norms are "bad" as such. Many norms will be better or worse than some alternative, and it is a crucial aspect of evaluating those norms to also identify those

[30] David Skarbek, *The Social Order of the Underworld: How Prison Gangs Govern the American Penal System* (New York: Oxford University Press, 2014).

[31] These answers are admittedly sketchy. For a more detailed analysis, see: Thrasher and Handfield, "Honor and Violence: An Account of Feuds, Duels, and Honor Killings," *Human Nature*, forthcoming 2018.

[32] Although, see Nicholas Southwood and David Wiens, "'Actual' Does Not Imply 'Feasible'," *Philosophical Studies* 173, no. 11 (2016): 3037–60.

alternatives and to judge their feasibility. If this is right, feasibility analysis and norm evaluation will be closely related. A "bad" norm with worse feasible alternatives or no feasible alternative may be bad in some sense, but as good as it gets in the current context. David Skarbek's analysis of prison populations helps us to understand that often what looks like irrational violence can have an underlying order. Often any organization of violence by norms is an important improvement over the alternative.[33]

To change these norms, as I argued with the honor killing cases, more than beliefs and values will need to change. Somehow the functional role that the norm currently performs will need to be performed by another norm or be rendered unnecessary. This can be done in a variety of ways, but understanding the beliefs and preferences of the individuals in the norm group as well as the historical/cultural context will be crucially important. Although it will be important to identify broad classes of norms, such as honor norms, and their functions, we should not be too confident in how much the analysis will translate from one place to another. Still, humans are pretty similar in their essential rationality and their need and desire to improve their lives and the lives of their families. We encounter many of the same problems in different guises, and although different in substance, it may be that the form of many norms share similar features and have similar functions.

One of the goals of this approach has been to avoid parochialism. Parochialism is a kind of defect, despite how common and natural it is. As people, we often forget the common humanity that binds us together. Parochialism can also blind us to what norms are actually doing in the societies where they exist. Developing and refining a nonparochial standard for evaluating norms is not only theoretically important, but it may also help to avoid approaches that may prove to be counterproductive in the "wild." Hopefully the comparative-functional approach and the integrative analytical framework in which it is embedded will help theorists as well as practitioners in this field.

*Philosophy, Chapman University*

[33] On this point see Douglass C. North, John Joseph Wallis, and Barry R. Weingast, *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History* (New York: Cambridge University Press, 2009).