

Rational reconstructions and the question of function

Rationality and Society
2019, Vol. 31(4) 409–431
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1043463119883959
journals.sagepub.com/home/rss



Lina Eriksson 

University of Gothenburg, Sweden

Abstract

Social norm emergence is commonly explained by stating that norms serve certain functions – for example, solving cooperation or coordination problems. But critics argue that examples of norms that do not seem to serve functions show that functions cannot explain social norms. However, both sides tend to make assumptions about how explanations of social norms in terms of functions would work. By discussing four problems for these assumptions, I will show that they are over-simplified. Instead of asking whether norms serve functions, we need to ask more specific questions about the relationship between the norm and the function it supposedly serves.

Keywords

Norm emergence, rational reconstruction, social norms

Introduction

A common view in the literature on social norms¹ is that norms emerge because they serve certain functions (e.g. solving cooperation or coordination problems; Arrow, 1971; Coleman, 1990; Cooter, 1996; Ellickson, 1991; Ullmann-Margalit, 1977; Voss, 2001). To outline such functions is to provide *rational reconstructions* of norms. Critics respond that many norms do not seem to serve functions, and so therefore cannot be explained in terms of such functions (Elster, 1989; Posner, 2000).

Corresponding author:

Lina Eriksson, University of Gothenburg, Box 100, 405 30 Gothenburg, Sweden.

Email: lina.eriksson@gu.se

Both sides of the debate tend to assume that the explanatory relationship between functions and norm emergence is a fairly simple one: either rational reconstructions of norm function have great explanatory value for answering questions about norm emergence, or functions are largely irrelevant for the explanation of norm emergence.

To support their respective views, they provide anecdotal evidence: rational reconstructivists offer examples of norms that seem to serve a function very well, and their critics offer examples of norms that do not seem to serve a function at all. Although rational reconstructivists and their critics obviously disagree about the role functions play in explaining norm emergence, they seem to share some fundamental assumptions about that role. These are as follows:

- *Meaning.* It is always clear what it means that a norm serves a function (including in particular that a norm serves a function efficiently).
- *Assessment.* It is always straightforward to determine whether a norm serves a function (including in particular that a norm serves a function efficiently).
- *Efficiency.* Norms either serve functions efficiently (in some sense) or not at all.

If, however, these assumptions are wrong, the explanatory relationship between functions and norm emergence can be much more complicated, and instead of arguing in favour of or against that functions have an explanatory role to play, we should ask a whole set of other questions about the explanations of norm emergence. I will indeed argue that the assumptions do not hold generally, and that the explanatory relationship between functions and norm emergence is more complicated than rational reconstructivists and their critics tend to acknowledge. Of course, it is possible – and indeed likely – that if asked, proponents of both sides will gladly acknowledge this. However, they do not tend to discuss the ways in which the assumptions are over-simplified, and thus the debate about the relationship between functions and explanations of norm emergence has continued to rely on these over-simplified assumptions.

As noted, I will argue that the overly simple assumptions (and thereby the tendency to use anecdotal evidence based on those assumptions) make us fail to ask a number of relevant questions about the relationship between functions and the explanation of norm emergence. The fact that anecdotal evidence can be insufficient evidence for or against functional explanations is of course a broader problem than what I can discuss here. For example, in

a paper on functional explanations in social science, Harold Kincaid (1990) makes the general point that using anecdotal evidence against a functional explanation can dismiss such explanations too quickly, because there is a number of ways in which the causal connections between a function and the phenomenon that supposedly exists because of that function, can be more complicated than the simple connection assumed in these anecdotes.² However, the kind of issues that I raise in relation to the explanation of norm emergence are slightly different from the ones Kincaid raises in more general terms, and I have to set such broader concerns aside.

In the following section, I will outline the rational reconstructivist approach to explanations of norm emergence, as well as the views of the critics. After that, I will discuss reasons why the assumptions are oversimplified. The first problem concerns for whom the norm is supposed to serve a function. It is not always clear what it means that a norm serves a function for somebody, or how to assess for whom it supposedly serves that function. The second problem concerns the assumption about efficiency: I will argue that serving a function is not an all-or-nothing affair, and therefore, a demonstration that a particular norm *N* does not serve a function efficiently does not show that it does not do so at all, or that the function had no role to play in the explanation of the norm's emergence. The third problem is that norms do not operate in isolation but always against a background of other norms that can restrict what problem solutions are feasible, and this can make it difficult to assess whether any single norm is efficient or not. Finally, I raise the point that sometimes there are more than one possible function that a given norm *N* might be thought to serve, and in those cases, it can be hard to determine which – if any – of the possible functions is *the* function that the norm primarily serves, and/or which explains its emergence.

Rational reconstructions

Standard accounts of what a social norm is emphasize that a social norm is a behavioural regularity – it concerns behaviour that is widespread. But they also point out that for a social norm to exist, it is not enough that most people behave this way, but they must also think that people *ought* to behave this way, and disapprove of them not behaving that way. Another crucial aspect is expectations: most people expect that most others will behave accordingly, and that others will expect them to behave this way, and think that they ought to do so. And finally, social norms come with a sanctioning system; if you do not comply with the social norm, you risk sanctions from others, ranging from disapproval to social ostracism or even violence.³

A rational reconstruction of a norm aims to show how compliance with the norm in question can be the equilibrium strategy for rational agents, and this is of course crucial for understanding norm compliance. If a norm cannot be so reconstructed, it would be unstable: if rational agents really would do better by not complying with the norm, once everything is said and done, then the real puzzle is why they comply. Of course, one might argue that real people are not rational agents. But the extent to which rationality is an idealization in rational reconstructivist arguments is quite limited and innocent: the incentives to which agents are supposedly sensitive are incentives we are all familiar with in our everyday lives.

Rational reconstructions are reconstructions of the function of norms in rational choice terms. They show how norms serve the interests of rational agents by solving some problem that these face. These reconstructions are usually game-theoretic (but not always; for a notable exception, see Posner, discussed below), depicting norms as solutions to games of either coordination or cooperation. Thus, they show how norms allow us to coordinate our actions, or to avoid sub-optimal outcomes in Prisoner's Dilemma games. Norms also make others' behaviour predictable, which can provide the required assurance in Assurance games.

The fact that norms can be very useful things – coordinating behaviour, facilitating cooperation, providing assurance and making others' behaviour predictable – has led many to suggest that norms exist *because* they are useful (see, for example, Arrow, 1971; Coleman, 1990; Cooter, 1996; Ellickson, 1991; Ullmann-Margalit, 1977; Voss, 2001). They exist because they solve certain problems we have, that is, because they serve certain *functions*. Consider, for example, Edna Ullmann-Margalit's account of what her rational reconstructions explain in the classic *The Emergence of Norms*:

My basic argument is that certain types of norms are solutions to problems posed by certain interaction situations. These problems inhere in the structure – in the game-theoretical sense of structure – of the situations concerned, pertaining to some or all of the interacting participants. (Ullmann-Margalit, 1977: 10)

It is crucial to note what it is that rational reconstructivists take themselves to be explaining: they explain, they claim, *why* norms emerge, not *how* they do so. The latter, it is claimed, can only be explained through historical analysis (Ullmann-Margalit, 1977). Similarly, James Coleman (1990) argues that the existence of coordination and cooperation problems generates a 'demand for norms' (see also Voss, 2001). Rational reconstructions are thus a kind of *functional* explanation. Norms tend to emerge *because they serve some function* (e.g. like ensuring cooperation), even if the existence of such a function is not a sufficient condition for the

emergence of a norm. The kind of functional explanation inherent in rational reconstructions of norms is however not functional explanation of the traditional kind, since traditional functional explanations tend to focus on the group interest. Rather, rational choice theory requires that it can be shown how cooperation can be an equilibrium strategy for each and every participant (or, to be more specific; for each and every rational participant with a typical utility function), not just why cooperation is beneficial for the group as a collective. But nevertheless, the *function* is what explains why the norm emerges.⁴

Generally speaking, a norm serves a function if it has positive consequences of some sort, and if these consequences are not accidental. But it is not entirely clear what it means that a norm has positive consequences. A common notion in the literature is that a norm has positive consequences when everybody benefits, but sometimes this is qualified somewhat, so that the claim is that the norm has positive consequences when almost everyone benefits (e.g. Elster (1989: 108) uses both meanings of 'positive consequences'). But it is also common to leave the exact meaning more vague, as when Arrow (1971) writes that 'it is useful for individuals to have some trust in each other's word' (p. 22). Presumably, he means either all individuals or almost all individuals. But since I will at some points in this article be discussing optimally good consequences, rather than just positive consequences, it is worth pointing out that none of these examples (and many others in the literature) discuss whether there could be some further improvement that would be Pareto-desirable. Maybe everyone benefits from a norm S_1 , but a different norm S_2 would still be a Pareto-improvement. Or perhaps, everyone benefits from S_1 and everyone would benefit from S_2 , but some benefit more from S_1 than from S_2 , and for others, it is the other way around. Since the literature usually is silent on such issues, I will leave the exact notions of positive consequences and efficiency similarly unspecified.

People's views on the merits of rational reconstructions tend to fall into two separate and opposite groups. Some support the functionalist story wholeheartedly, some with more elaborate accounts or arguments than others. For example, Robert Ellickson (1991, 2001) argues that norms tend to be efficient solutions to problems, although he in later work recognizes that norms can be somewhat inefficient solutions. Kenneth Arrow (1971) considers (at least some) social norms to be agreements that solve market failures, thereby improving the efficiency of the economic system (p. 22). Cooter (1996) argues that norms will arise to create public goods, when contributions to those public goods simultaneously allow agents to signal their 'good type'. Norms thus help mitigate the undersupply of public goods. And norms with negative consequences are rare, he claims, since in

most cases, in which the signalling behaviour does not also involve socially desirable behaviour, no norm emerges (Cooter, 1996: 1677–1678). When explaining actually existing norms, rational reconstructivists thus tend to look for a function that the norm serves. For example, although the norm that men fight duels to resolve conflicts is usually considered a textbook example of a norm that has negative rather than positive consequences, Schwartz et al. (1984) argue that the norms surrounding the duels fought by gentlemen in the Southern United States were not as inefficient as they may first seem. The norms served several functions at once. One important function was the revelation either of a person's honour or of likely future behaviour. But the main function, Schwartz et al. seem to claim, was that of preserving a socially tight-knit community in which nobody mistreated slaves (over and above owning slaves itself), since such mistreatment would raise the risk of a slave rebellion. However, their paper mostly discusses the extent to which the benefits of revealing character and trustworthiness could outweigh the obvious disadvantages of a practice that involved the risk of death for important members of society, and the result of which had little correlation with who had originally been in the wrong.⁵

As noted in the 'Introduction' section, this type of functionalist explanation is the target of a standard argument against rational reconstructions of norms. Jon Elster (1989) argues that the functionalist story must be wrong since not all norms have beneficial consequences; some of them even have really negative consequences, for example, norms of revenge (pp. 107–113). Elster is not the only one who takes this view. For example, Eric Posner (2000), in his work on the relation between social norms and the law, dismisses functionalism with a quick wave of the hand: '(F)unctionalism . . . is empirically false and methodologically sterile' (p. 172). Posner also explicitly refers to Elster's argument as the reason for his rejection of functionalism. Others have noted their scepticism in milder words, agreeing with Elster and Posner that uncritical functional explanations are implausible, but nevertheless arguing that functions do have some explanatory role to play (see, for example, Bicchieri and Muldoon, 2011).

However, an over-simplified view of the relationship between functions and social norm emergence means that there are aspects of that relationship that we do not investigate. In the following section, I will discuss such over-simplifications, and provide a more nuanced view of the relationship in question.

Function for whom – and when?

In assessing whether a norm serves a function, we obviously need to be able to point out for whom it serves such a function. But in this section, I will

argue that it can be far from straightforward to determine which group, if any, a given norm serves a function for. Note, though, that with ‘function for a group’, I do not mean to imply that there is some sort of mysterious group interest, only that the norm serves a function for each individual in this group (with some qualifications that I will discuss below). A difficulty to identify the group of individuals the norm serves a function for is obviously a problem for the rational reconstructivists. But it is also a problem for their critics, since justifying a claim that a norm does not serve a function requires that we can rule out that there is a relevant group for which it does so. Many of the rational reconstructivists’ main examples (norms against littering, against freeriding on others’ generosity, against not doing your fair share, etc.) portray norms as solving general problems in ways that serve all agents equally well. But of course, societies are full of norms that serve the interests of some groups but not others. For example, the norm that members of criminal organizations do not speak to the police serves the interests of the criminal organizations, but not of general society. However, there is nothing about rational reconstructions that necessarily makes it the case that the functions they discuss have to be functions for the whole of society – and indeed, many rational reconstruction examples make this perfectly clear.

However, it is not enough to recognize that a norm does not have to be beneficial for the whole of society. If it was, all we would have to do to get things right would be to focus on the smaller group of people who are complying with the norm (as in members of criminal organizations). And, as we will see, although this might work in the case of the criminal organization, it is not always the case that the group of agents who are currently complying with the norm is also the group of individuals each of whom benefits from the norm.

Take instead an example of a norm that requires inequality. For example, consider a norm that requires that men get to eat before women, so that the former can eat their fill and choose the best pieces of food, whereas the latter eat what is left. The norm obviously applies to both men and women, in that it prescribes behaviour for both men and women. In this sense, it is not like the norm that members of a criminal organization do not speak to the police, as this norm does not prescribe any behaviour in particular to people who are not members of criminal organizations. But it is obviously not plausible that the sexist norm should be explained in terms of a function it serves for *both* men and women. In general, when a norm applies to a group, but only some members of that group benefit from the norm, an explanation of the emergence of that norm in terms of function will not be about a function it serves for *each* member of the group, but only for the dominant members.

Some social scientists and philosophers use game theory to understand bargaining between agents with conflicting interests and with unequal bargaining power (see, for example, Knight, 1992). Obviously, we can apply the same approach to our study of norms, and few if any rational reconstructivists would disagree. But before we therefore conclude that the fact that costs and benefits from a particular norm are unequally distributed across agents does not pose a problem for rational reconstructivists, we should recognize that in practice, the unequal distribution of costs and benefits can make it very difficult to establish whether a function *F* explains the emergence of a norm *N*. Identifying who, in the group of people who take themselves to be expected to comply with the norm, benefits from it can be difficult. Taking care not to ascribe an explanatory role to some function *F* unless we can identify for whom the norm in question served function *F*, thus means that we must pay attention to important individual or sub-group variation in costs and benefits.⁶

The really difficult situation for anyone wanting to claim that a function *F* explained the emergence of a norm *N* is however a different one, namely one of time. It may be the case that *nobody* in the current group of norm compliers benefits very much from the norm. Suppose that the norm was originally created by an earlier group of agents. It is possible that the norm emerged because it served the interests of (some of) the members of *that* group. But norms have a tendency to take on a life of their own, once they exist.⁷ Because of possible coordination effects and the threat of sanctions against violators, people may very well have an incentive to comply with the norm now, even though the problem that the norm emerged to solve is long since gone. This is similar to a problem for adaptationist explanations in biology: a particular trait might have arisen as an adaptation once upon a time, but might no longer be adaptive and instead might persist for some other reason. Just as the reason why a biological trait still exists might be different from the reason it once emerged, the reason why the norm continues to exist may not be the same as the reason why the norm first emerged. Of course, it is possible that function is a better explanation of norm persistence than of norm emergence. But that question is not the subject of this article. Similarly, a biological trait could look like an adaptation to some aspect of the current environment, but still not have arisen as an adaptation. Issues like these are part of what makes adaptationist explanations difficult in evolutionary biology.

Although not theoretically problematic for rational reconstructivists or for their critics, the problem of when, more exactly, a norm served a function (or not, as it may be) for a relevant group of individuals, is empirically difficult to deal with. The kind of anecdotal evidence provided by rational

reconstructivists is not enough to establish their claim that the norms in their examples serve functions. It may very well be that a norm *N* happens to serve a function *F* for some group *G* *now*, but to establish that the norm emerged because of the function we would have to establish that the norm served that function when the norm emerged – and if we recognize that whether a norm serves a function can change over time, we can no longer jump to conclusions about this from the mere fact that it serves a function *now*. The function it serves now might, as noted above, merely explain why the norm persists.⁸ The question of why a norm persists is different from the question of why it emerged; sometimes, the two questions have the same answer, but sometimes they do not.

But although the problem that we cannot easily draw a conclusion about what function a norm served when it emerged from observations about what function it serves now, obviously is a problem for rational reconstructivists, it can be equally problematic for critics of rational reconstructions. Although it may be true that some norm they discuss do not serve any function (or at least do not do so efficiently, see the next section) *now*, for any group, this does not preclude that some function explained why the norm originally emerged. This point about the problems of establishing – or denying – that a norm emerged because it served a function becomes even more difficult to deal with empirically once we take into account that although the norm might serve a function, at a given point in time and for some group or sub-group, the power relationship between various sub-groups and groups can change over time, in ways that could mean that although the norm *N* might serve a function *F* for a group *G* now, and there existed a group *G**, with similar interests as that of *G*, when the norm *N* emerged, it may still not be the case that *G** was powerful enough that the norm *N* emerged because it would serve *G**'s interests. A good argument about how a particular norm *N* emerged because it served a function *F* would thus have to be much more careful about for whom and when the norm *N* supposedly served the function *F* than what is currently common.

Rather than speaking of a norm that serves a function, we should thus speak of a norm *N* that serves a function *F* for the individuals in group *G* at time t_i .

Efficient and inefficient norms

Let us now turn to the assumption of norm efficiency. Although Elster (1989) lists not just a social norm of revenge but also several other norms as counterexamples against the claim that norms exist because they have good collective benefits (pp. 107–113), the example he discussed in most detail

(in Elster 1990) is the norm of revenge. Elster discusses the norm of revenge from the point of view both of individual and collective benefits, and denies that the norm have either type of benefits. A rational reconstructivist explanation of the norm of revenge would claim that the norm in question serves the function of maintaining peace, which is important for all (or almost all) individuals in the group, and further that it is individually rational to comply with the norm to realize this benefit. If Elster were right, a rational reconstructivist explanation of the norm of revenge would thus fail.

Revenge, in particular *vendetta*, norms say that if you are wronged (be it the murder of a family member or an insult in the bus queue), you have to take revenge. The revenge has to be severe enough that your honour is restored: quite often, this involves causing more damage to the other party than he or she caused you. Since they, too, accept the revenge norm, however, they will then have to take revenge on *you* for the damage you caused, and the two of you will cause greater and greater damage until, in the extreme case, eventually one of you is dead (and it might not even end then). Of course, things do not have to spiral out of control; maybe the offending party decides to accept your revenge and let it be. But even so, the damage caused to them will have been greater than the damage done to you. All in all, norms of revenge are thus rather destructive. Therefore, Elster (1990) argues, norms such as norms of revenge cannot have emerged as solutions to a problem. Nevertheless, we do know that people often punish those that wrong them,⁹ and that sometimes punishing wrongs is mandated by a social norm.

Note that the argument against an explanation of norm emergence in terms of functions does not require an example of a norm that has *negative* consequences, only of a norm that does not seem to serve any function. There are two types of norms that do not serve any function: those with negative consequences, of course, and also those that simply have no effect on people's well-being. The argument that not all norms serve functions can be supported by an example of a norm of either of these two types – one with no effect on well-being, or one that has negative effects. Elster makes his point through the example of an outright destructive norm, one that has a negative rather than a neutral impact. A possible explanation for his choice of example is that because of its negative effects, the *vendetta* norm is such a striking example of a norm that does not seem to benefit those that comply with it.

It is also worth pointing out that there are two ways in which a norm can fail to have beneficial consequences. One, which I have stressed above, is that the norm is not an efficient solution to the problem it has emerged to solve. The failure thus consists in inefficiency. The other way in which a

norm can fail to be beneficial is that it has negative consequences. It is worth noting that a norm could, in principle, be an efficient solution to a problem, and yet have negative side effects. For example, a norm that forbids women to move unescorted in public is quite efficient in solving the problem of the risk for women of being raped by a stranger in a public place. But it does so at a cost that is so high that women are better off without the norm in question. Similarly, one could imagine a very inefficient norm that nevertheless does not do much damage. When Elster argues that the vendetta norm is an example that shows that norms do not serve functions, he seems to be alluding to both ways a norm can fail to be beneficial: his argument is that since there are instances of serious violence, the vendetta norm does not uphold the peace (a claim about inefficiency), and that these instances of serious violence are destructive (a claim about negative consequences). My main claim in this article concerns the efficiency, or lack thereof, of norms, since this is the aspect of the lack of beneficial consequences that is most obviously connected to the question of what role functions play for the emergence of norms.

How could a defender of the explanatory importance of functions answer Elster? One option would be to argue that for each person, the vendetta norm serves the function of helping them to overcome time inconsistency. Since actually retaliating against an insult is costly, a rational agent would not follow through on any threats made to retaliate, were the insult made despite those threats. But since everyone would know this, no threats to retaliate would be credible. It might overall be better to be the kind of person who is irrational enough to retaliate, and known to be so, since people would avoid provoking such a person (Frank, 1988). But how can a rational person become such a person? The honour system makes it possible: the norm to take revenge raises the cost of abstaining from retaliation, so that even rational agents will find it in their interests to retaliate, and the norm also makes it generally known that insults are typically revenged. The result is that fewer insults are made. However, Elster (1989) argues that this line of reasoning is mistaken. He points out that people with strong honour codes only benefit in environments in which most others do not comply with such codes: 'Mafiosi seem to do better for themselves in the United States than in Sicily' (p. 106). Therefore, when a majority complies with honour code norms, it is no longer beneficial to be the kind of person who takes revenge against anyone who insults them. Elster also argues that it is indeed characteristic of norms that they are complied with even when such compliance is not in the interest of the person who complies: we take revenge because of what was done to us, not because of some outcome we want to achieve. It is worth noting, though, as Hamlin (1991) does, that even

though having been wronged might be a *necessary* condition for taking revenge, it need not be a sufficient one. Indeed, Hamlin argues, if a social norm is defined partly through a system of approval and disapproval for relevant actions, taking revenge cannot be a social norm unless it is connected to honour, because without the context of honour, we have no reason to approve or disapprove of others' vengefulness (or lack thereof), unless it interferes with our own pursuits. Actions taken to protect one's honour, on the other hand, are subject to approval and disapproval. However, actions performed to gain approval or avoid disapproval are actions performed for the sake of an outcome we want to achieve, not only because of what was done to us. Furthermore, Hamlin also points out that if we treat honour as a form of utility itself, taking revenge can be perfectly rational, and since the notion that people are partly motivated by concerns of honour is hardly an ad hoc assumption, interpreting revenge as attempts to restore or protect honour makes sense.

According to another account of norms with a negative or neutral effect on well-being, such norms are sustained because people mistakenly believe that everyone else, or almost everyone else, approve of the norm, even though in fact nobody or almost nobody does. Social norms are sustained by the kind of common attitudes that satisfy the conditions that (almost) every member of the group holds the attitude in question, believes that (almost) every other member of the group holds the attitude in question, believes that (almost) every other member of the group believes that (almost) every other member of the group holds the attitude, and so on (for slight variations in these conditions, see Lewis, 1969; List, 2014). But as Kuran (1995), Prentice and Miller (1996) and List (2014) have argued, if in fact nobody or almost nobody held the attitude in question, but the other conditions are satisfied (people believe that every or almost every other member holds the attitude, etc.), there can still be what Prentice and Miller refers to as *pluralistic ignorance*, where what List refers to as a *falsely attributed common attitude* can sustain a social norm, even if the norm is in fact deeply unpopular.

Furthermore, Posner's account of norm compliance as signalling suggests yet another possible explanation of norms that do not have positive effects; if norm compliance is simply a matter of signalling that one has a low discount rate for future benefits, then it does not matter whether the norm in question serves some other function. People with low discount rates will comply as long as compliance signals low discount rates. However, Druzin (2011, 2013), further developing Posner's account, argues that norms that serve some (other) function than merely signalling will be salient forms of behaviour that allow for the kind of coordination around shared understandings of the symbolic meaning of behaviour that is necessary for

any form of behaviour to signal anything. Just as some trait can be co-opted to fulfil a(nother) function than it originally did during the evolution of a species, so a functional norm can be co-opted into serving as a signal.

However, I want to draw attention to another problem with Elster's social norm of revenge example, namely that this norm does in fact serve a function: it upholds peace. Given the norm of revenge, peaceful interactions are indeed the Nash equilibrium; if everyone knows that everyone will retaliate, and with more force than the original offence, then everyone will abstain from being offensive. This, it has actually been suggested, is the explanation for why Southern United States has a culture of friendliness and politeness. Since the Southern culture includes the norm of revenge, you had better be nice for any unpleasantness will trigger serious retaliation. Given that all others play the strategy, 'be nice if others are nice, take revenge if others insult you', you had better do the same. The result is peace (Cohen et al., 1996).

Elster argues that it is not clear that it is good that there are fewer quarrels in societies with honour codes than in societies without. The cost of suppressing these quarrels may be too high. But notice that if Elster is right that the cost of preventing quarrels is higher than the cost such quarrels generate, this only means that we might be better off with the quarrels than we would were we to pay the cost of upholding peace. This does not mean that the norm of revenge does not serve the function of upholding peace.

However, we should be more specific about the peace that results from the revenge norm: there is peace *until someone makes a mistake*. Maybe they are drunk and fail to consider the consequences, maybe they think they will turn out to be stronger and tougher than the other (in fact, they might be right). Nash equilibria are, unfortunately, not necessarily robust, and a mistake can push play off the equilibrium path. When this happens, the nasty side of the norm is triggered. Given this, maybe a norm of revenge is not *the best possible* solution to the problem of how to ensure peaceful interactions, because it is a rather risky one. But the fact that a norm of revenge is not the *best possible* solution to the problem does not mean that it does not serve the function of upholding peace at all.

Elster's mistake is thinking that serving a function is necessarily an all-or-nothing affair. It is not. Pointing out that not all norms serve functions really efficiently is not enough to demonstrate that functions have nothing to do with the explanation of the emergence of these norms. Norms that serve functions can do so to different degrees: it may be that although they do serve the function to some extent, they do so rather inefficiently. In particular, they may not serve the function as well as an alternative norm would have done. There are lots of possible reasons why a given norm might not

be the *most* efficient possible solution to a problem even though being a solution to that problem is an essential feature of the norm in question (e.g. path dependency, collective action problems, lack of information). Elster thus drew a stronger conclusion than what was warranted by his example of vendetta norms: demonstrating that there are norms that do not serve functions really efficiently does not prove that functions have nothing to do with the explanation of the emergence of these norms. Instead of speaking of a norm N serving – or not serving – a function F , we should therefore say that the norm N serves the function F for the individuals in group G at time t_i to a degree D , which ranges from zero to one, inclusive.

However, it is worth noting that the problem remains of how to establish that a function F , served to some degree D by a norm N , explains why the norm N emerged. As Brennan et al. point out, unless we know the mechanism through which a norm has emerged, and in particular, know that this mechanism was indeed selecting on the basis of the function in question, we cannot say that that function explains *why* the norm has emerged. The problem exists because as long as we accept functions that a given norm serves to some degree but not very efficiently, we can always find *some* function that the norm serves. Often we can find several (e.g. see the discussion below about etiquette rules). So how can we distinguish between cases in which a particular, given function explains *why* the norm emerged from cases in which the fact that the norm served that function to some degree was just a happy coincidence? Indeed, the literature on functions in biology have long acknowledged that sometimes a trait can have accidental benefits, without thereby having developed to generate those particular benefits (see, for example, Kincaid, 1990, and the distinction between an explanatory and a descriptive use of the term ‘function’ in Mitchell (1989)). Brennan et al. argue that when a norm serves a single function and does so really efficiently, it might be reasonable to assume that that function explains the emergence of the norm. However, when a norm does not serve any function very efficiently and/or serve more than one function to some degree, the major explanatory work is done by information about the mechanism: the function(s) might or might not be a coincidence (Brennan et al., 2013).

The importance of other norms

Consider again the norm of revenge that Elster argued did not serve a function, and in particular, what happens when behaviour has indeed been pushed away from the equilibrium position. In a society characterized by the vendetta norm, behaviour is likely to become increasingly violent, thus

pushing behaviour further and further away from the previous equilibrium state. However, if there were forces that instead push behaviour back towards the equilibrium state after a small deviation, peace would be quickly restored without the negative consequences of the vendetta norm. Other norms can be such a force. Consider the norm against causing a scene, which is quite strong in the United Kingdom, Sweden and many other countries. Brits and Swedes do get angry and may take violent revenge, but they would generally have suffered major and usually repeated offences before that would happen. There is obviously no vendetta norm in these societies, so we could say that their revenge norm was simply *different*. Although that is certainly correct, it also seems that the norm against causing a scene, which applies more generally than just to cases of potential revenge, is important for making the revenge norm in these societies what it is. The stability of the peace equilibrium affects the extent to which the norm succeeds in upholding peace, but that stability in turn depends partly on what other norms there are. But if the stability of a particular norm equilibrium, in the sense discussed here, depends partly on what *other* norms exist in the relevant community, then assessing whether, and if so to what extent, a given norm serves a function will have to take into account whether the norm is, indeed, perhaps the most efficient norm *given these other norms*. It is thus not possible to assess just the extent to which *this* norm serves its function, without taking into account how other norms affect the situation.¹⁰ The argument is somewhat analogous to the argument in evolutionary biology that even if a trait is an adaptation, it will be a local optimum, not necessarily a global one: what traits an organism can develop depends crucially on what traits the organism already has. Similarly, what norm can serve a certain function depends crucially on what other norms already exist.

The point about the importance of the wider web of norms can be made more generally. The literature on rational reconstructions tends to treat norms as easily identifiable and separable ‘entities’, the functions of which can be determined independently of each other. But this is not how most norms work. Just as Sunstein (1997) remarks that it does not make sense to attempt to examine people’s interests as if they lived in a norm-free environment (ch 2), it makes little sense to examine a particular norm – to ask questions about why it has emerged, whether it serves a function and so on – as if it existed in an otherwise norm-free environment. Consider again the norm of revenge. It is embedded in a web of honour code norms. Therefore, it does not make sense to ask whether taking revenge serves a function, without also taking into account what behaviour these *other* norms require. The reason why it is important to take revenge against an insult is tied up with the importance given to protecting one’s honour in other types of

situations. As Cohen and Vandello write (1998) about the honour code related violence in the Southern United States,

[S]outhern violence is a product of a coherent, complete meaning system that defines the self, insults, honor, rituals for conflict, and the tools and methods—occasionally including violence—that are to be used when order is disrupted. (p. 584)

In honour cultures, having a reputation for being an ‘honourable man’ is essential for being thought to be trustworthy and to have integrity and courage, and notions of honour are built into the concept of masculinity. A real man responds to insults with violence, and to violence with more violence. The norm of revenge is just one norm surrounding the value of honour in these cultures. It is therefore a mistake to analyse the norm of revenge separately. It is of course possible that some of these honour norms can be replaced with new norms, while the rest of the honour norms are left intact. But in other cases, the other honour norms severely constrain what norms are possible, so that the particular honour norm we are concerned with cannot be easily replaced. This may very well be the case for the norm of revenge. Consider what would happen if the norms governing masculinity were different: if the appropriate way for a man to react to an insult was to ignore such childish displays of bad temper, it would be hard to see how there could be a norm of revenge that prescribed risky, violent retaliation. Since the norms about masculinity do *not* prescribe calm and indifference in the societies in question, but risky and violent behaviour, then it is similarly hard to see how there could not be a norm of revenge much like the one that Elster discusses. Therefore, we should recognize the possibility that although the norm of revenge has some negative consequences, we cannot rule out that it might still be the most efficient peace-keeping norm to have *given* the existence of this specific honour culture, since a different revenge norm would conflict with the rest of the honour norms and therefore be hard to sustain.

Elster indirectly acknowledges that the norm of revenge is but one norm in a package of honour norms: it is not clear, he claims, that there really is less violence in a society with a norm of revenge than in a state of nature, if we assume that in the latter, people are rational, because rational people would not harm each other ‘just to get even’, and honour codes generate fights because people want to show courage and ‘brinkmanship’. ‘The net effect’, he says, ‘is anybody’s guess’ (Elster, 1989: 112). ‘Getting even’ is obviously part of the norm of revenge itself. But the honour codes that generate fights for men to demonstrate their courage and masculinity are not parts of the norm of revenge. Rather, they are other honour norms, which

happen to interact with the norm of revenge in a particularly destructive way.

Instead of assessing the degree D to which a particular norm N serves a function F for group G at time t_i , then, we should assess the degree D to which a particular norm N serves a function F for the individuals in group G at time t_i and in norm context NC .

The problem of multiple possible functions for a given norm

Finally, I want to draw attention to a related problem that arises from the ones already discussed. This is the problem that sometimes it is not even straightforward to identify which function, if any, that a given norm supposedly serves. For example, consider the function of duelling norms. Ellickson (2001) claims that the old norm in the American South that gentlemen defend their honour by duelling served the function of maintaining solidarity among the white elite as a defence against slave rebellions. This interpretation is also found in Schwartz et al. (1984). To avoid slave rebellions, it was necessary to have – and uphold – a strong social norm against abusing one's slaves 'too much'. The argument is not spelled out in detail, but presumably, the thought is that if abuse went against an understanding of 'honourable behaviour', and honourable behaviour in general was maintained by duels, then duels served an important function because of the importance of avoiding abuse. Notice, however, that the duelling norm is considered as part of and embedded within a web of honour code norms. It is this web of honour code norms that serves the function of suppressing 'excessive' abuse of slaves. However, in contrast, Cohen and Vandello (1998) argue that the honour culture stems from the historical origins of herding societies: the South, and later also the West, of the United States were originally herding societies, where people lived far apart, law enforcement was scarce and unreliable, and cattle could easily be taken: it was therefore important to establish a reputation for being someone you did not want to 'mess with'. Presumably, there are ways to settle the disagreement about the origins of the honour culture and the associated duelling norm in the Southern United States. But it is obviously not easy, and part of the problem is that the function we ascribe to the duelling norm is dependent on the function we ascribe to the honour culture of which it was a part.

In other cases, it is even more difficult to assess which function, if any, is actually being served. Consider, for example, the case of etiquette rules. It would be absurd to argue that some important general welfare goal is served by people chewing with their mouths closed. However, keeping your mouth

closed when chewing is in many groups a social norm. Presumably, the explanation for this norm is tied to the explanation for etiquette rules in general. However, it is not clear which function, if any, that etiquette rules serve, and thus not clear which function, if any, that is served by the norm to chew with your mouth closed. The explanations of etiquette rules in terms of functions have included that etiquette rules are the upper class way of creating group identity, that etiquette rules help identify who belongs and who is an 'upstart' (Bourdieu, 1979), that they allow us to signal that we are 'good types' who respect rules (Posner, 2000), that etiquette rules allow us to signal general respect for norms (in virtue of the fact that they seem to have no other function?) and that they allow us to practice norm compliance and sanctioning (Young, 2008). But which one is *the* function, and how do we tell? Or does the norm serve several functions at the same time, all of which were important for its emergence? Notice also that what function a norm *N* serves can change over time, so that even if we can pinpoint which function the norm serves *now*, another function might explain why it emerged.

To sum up: if we suppose that it is straightforward to identify for whom – and when – a norm supposedly serves a function, that norms either serve functions very efficiently or not at all, and so on, it seems it would be straightforward also to determine which function the norm in question serves. But if we recognize that it can be difficult to assess for whom a norm serves or served a function, that norms can serve functions to different degrees, and that judgements about efficiency often involves judgements about the existence of and fit with other norms, we should not expect the result to always be a clear identification of a (single) function that is being served by the norm in question.

Conclusion

The debate about whether norm emergence can be explained in terms of functions that norms serve tends to be too simplistic on both sides. It is either assumed that functions play a clear and large role in explaining norm emergence, or that they do not play a role at all. That assumption is in turn based on three other assumptions: the assumptions about Meaning, Assessment and Efficiency. A few stereotypical cases, in which it really is quite clear and straightforward, and in which the norm in question either serves its supposed function efficiently, or not at all, are taken to be representative of the great majority of cases. But in fact, in most cases, things are not so clear, and if we are going to develop better theories of norm emergence we need to recognize factors that complicate the relationship between functions and explanations of norm emergence. In this article, I have drawn

attention to some such factors. I have argued that claims about functions need to include a more sophisticated analysis of for whom the norms are supposedly functional, and when norms are supposed to be so, and that it is a mistake to think that norms either serve functions very efficiently or not at all. When we make claims about the efficiency of a norm, we should also consider not just the norm in isolation, but also take into account other norms to which it is connected in various ways. Finally, partly as a result of these issues, it is also sometimes unclear *which* function a particular norm supposedly serves. Rather than speaking of norms that serve functions, we should therefore speak of norms that serve/served a function F for the individuals in group G at time t_i to degree D in norm context NC.

Against this background, we could then ask even further questions. For example, if a norm emerges because of a demand for it (i.e. because it serves some function), how can we establish whose demand was the relevant demand that contributed to the emergence of the norm and when it did so? And what factors can affect whether a norm continues to exist even when a previous demand for it has long since disappeared, that is, when the norm *emerged* because it served a function at that time, but *persists* for other reasons? It is also possible that a norm that *emerged* for reasons that had nothing to do with functions, *persists* because once the norm existed, it *came to* serve a function.¹¹ Since emergence and persistence can be caused by different factors, studying a norm's function (or lack thereof) at the present time might be an unreliable method for learning about the reasons for its emergence.

Furthermore, we should ask questions not only about the degree to which a norm serves a function but also about how the explanatory relationship between functions and norm emergence is affected if the norm in question serves a function but not very efficiently. If a number of different norms all would serve a particular function to some extent, we should ask questions about why we have one rather than the other. For example, both the vendetta norm and the norm that requires insults to be ignored as childish outbursts of temper serve the function of upholding peace. But why do some cultures have one norm and other cultures the other?

And for a final example, when we recognize that the particular norm we are interested in is embedded in a web of other norms, it becomes clear that we need to ask to what extent a different norm would have been possible given this wider web of norms.

Some of the issues raised above regarding the difficulty of inferring from an outcome what the intended goal was, to establish a causal link between intentions and outcome, apply not only at the level of groups, but also at the individual level. When an individual acts in a way that happens to result in an outcome that is quite optimal given what we know or assume

about the individual's goals, the rationality assumption leads us to the claim that the individual in question acted *to* achieve that outcome. But when the outcome only relates to those goals in some rather minimal way, we hesitate to make such claims. Indeed, it is difficult to know whether we were wrong about the individual's goals, about the circumstances that affected her options or simply about her level of rationality. But the step from individual rationality to group rationality – such as the claim that social norms serve a particular function – complicates such inferences even more. And that should not be surprising. After all, rational reconstructions of norms are built on the notion of individual rationality, not necessarily of group rationality, and a commitment to the first does not imply a commitment to the latter.¹²

Acknowledgements

This work was supported by The Swedish Research Council.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: The paper is part of a project on social norm change funded by The Swedish Research Council, funding received January 2017, grant number 2016-02318.

ORCID iD

Lina Eriksson  <https://orcid.org/0000-0003-3205-6239>

Notes

1. From now on, I will for the sake of brevity use the expression 'norm' instead of 'social norm'. See the beginning of the section 'Rational reconstructions' for a characterization of social norms.
2. Kincaid mainly discusses persistence rather than emergence, but his argument applies to emergence as well.
3. For example, see Bicchieri (2006). The account outlined here is sufficiently detailed for my purposes. Furthermore, I also set aside the problem about how to distinguish social norms from, say, moral norms.
4. Note that one could give a rational reconstruction explanation of why a norm persists, without thereby claiming that the function alluded to explains why the norm emerged. Maybe some of the problems I will raise for rational reconstructions of norm emergence would not arise – or at least not to such an extent – for rational reconstructions of norm persistence. But the view I am discussing concerns norm emergence, not norm persistence.

5. The literature on norms, including the duelling norm, often does not distinguish between explanations of norm emergence and norm persistence.
6. For one example, see Voss (2001: 118) on individual differences in discount rates.
7. Just as institutions in general do; for a recent discussion of path dependency of formal and informal institutions, see Lecours (2005).
8. The problem is similar to that facing those who try to explain a current trait of a species as a genetic adaptation. To prove that a trait emerged because it increased reproductive success, we would have to prove that the trait did, in fact, increase reproductive success at the time of the trait's emergence. It is not enough that it increases reproductive success now (although having a positive effect on reproductive success now might explain why the trait persists). See also Kincaid (1990).
9. There is a large literature on people's tendency to punish wrong doers (see, for example, Fehr and Gächter, 2002). The tendency to punish seems to be general, but the extent of punishment varies culturally (see Henrich et al., 2006).
10. The claim is similar to the claim in biology that what traits will increase reproductive success depends partly on what other traits that individual already has. We may therefore achieve local, but not necessarily global, optimal states: reach our greatest potential reproductive fitness given what we are, rather than reach the greatest reproductive fitness that someone who was not like us could reach.
11. For a discussion of various ways in which the causal link between functions and persistence (and some of these apply to emergence as well) can be structured, see Kincaid (1990).
12. I am grateful to an anonymous reviewer for raising this point.

References

- Arrow K (1971) Political and economic evaluation of social effects and externalities. In: Intriligator M (ed.) *Frontiers of Quantitative Economics*. Amsterdam: North-Holland, pp. 3–25.
- Bicchieri C (2006) *The Grammar of Society*. New York: Cambridge University Press.
- Bicchieri C and Muldoon R (2011) Social norms. In: Zalta EN (ed.) *The Stanford Encyclopedia of Philosophy*. Available at: <http://plato.stanford.edu/archives/spr2011/entries/social-norms/>
- Bourdieu P (1979) *La Distinction*. Paris: Editions de Minuit.
- Brennan G, Eriksson L, Goodin R, et al. (2013) *Explaining Norms*. Oxford: Oxford University Press.
- Cohen D and Vandello J (1998) Meanings of violence. *The Journal of Legal Studies* 27: 567–584.
- Cohen D, Nisbett R, Bowdle B, et al. (1996) Insult, aggression, and the southern culture of honor: an 'experimental ethnography'. *Journal of Personality and Social Psychology* 70(5): 954–960.

- Coleman J (1990) *Foundations of Social Theory*. Cambridge, MA; London: The Belknap Press of Harvard University Press.
- Cooter D (1996) Decentralized law for a complex economy: the structural approach to adjudicating the new law merchant. *University of Pennsylvania Law Review* 144: 1643–1696.
- Druzin B (2011) Law, selfishness, and signals: an expansion of Posner's signaling theory of social norms. *Canadian Journal of Law & Jurisprudence* 24(1): 5–53.
- Druzin B (2013) Eating peas with one's fingers: a semiotic approach to law and social norms. *International Journal for the Semiotics of Law* 26: 257–274.
- Ellickson R (1991) *Order Without Law: How Neighbors Settle Disputes*. Cambridge, MA: Harvard University Press.
- Ellickson R (2001) The market for social norms. *American Law and Economics Review* 3(1): 1–49.
- Elster J (1989) Social norms and economic theory. *Journal of Economic Perspectives* 3(4): 99–117.
- Elster J (1990) Norms of revenge. *Ethics* 100: 862–885.
- Fehr E and Gächter S (2002) Altruistic punishment in humans. *Nature* 415: 137–140.
- Frank R (1988) *Passions Within Reason: The Strategic Role of Emotions*. New York: W. W. Norton.
- Hamlin A (1991) Rational revenge. *Ethics* 101(2): 374–381.
- Henrich J, McElreath R, Barr A, et al. (2006) Costly punishment across human societies. *Science* 312(5781): 1767–1770.
- Kincaid H (1990) Assessing functional explanations in the social sciences. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1: 341–354.
- Knight J (1992) *Institutions and Social Conflict*. Cambridge: Cambridge University Press.
- Kuran T (1995) *Private Truths, Public Lies*. Cambridge, MA: Harvard University Press.
- Lecours A (ed.) (2005) *New Institutionalism: Theory and Analysis*. Toronto, ON, Canada; Buffalo, NY; London: University of Toronto Press.
- Lewis D (1969) *Convention*. Cambridge, MA: Harvard University Press.
- List C (2014) Three kinds of collective attitudes. *Erkenntnis* 79(9): 1601–1622.
- Mitchell S (1989) The causal background of functional explanation. *International Studies in the Philosophy of Science* 3(2): 213–229.
- Posner E (2000) *Law and Social Norms*. Cambridge, MA: Harvard University Press.
- Prentice DA and Miller DT (1996) Pluralistic ignorance and the perpetuation of social norms by unwitting actors. *Advances in Experimental Social Psychology* 28: 161–209.
- Schwartz WF, Baker K and Ryan D (1984) The duel: can these gentlemen be acting efficiently? *The Journal of Legal Studies* 13: 321–355.
- Sunstein C (1997) *Free Markets and Social Justice*. New York: Oxford University Press.

- Ullmann-Margalit E (1977) *The Emergence of Norms*. Oxford: Oxford University Press.
- Voss T (2001) Game-theoretical perspectives on the emergence of social norms. In: Hechter M and Opp KD (eds) *Social Norms*. New York: Russell Sage Foundation, pp. 105–136.
- Young P (2008) Social norms. In: Durlauf S and Blume L (eds) *The New Palgrave Dictionary of Economics*. 2nd ed. London: Macmillan.