Social Theory and Practice
Vol. 45, No. 4 (October 2019): 579–599
DOI: 10.5840/soctheorpract2019123072

Social Norms as Signals

Lina Eriksson

Abstract: According to the signaling theory of social norms, people comply with social norms in order to signal that they have a low discount rate for future costs and benefits and thereby that they are reliable cooperation partners. But the theory does not take into sufficient account the fact that the signaling value of social norm compliance depends on how many other people that comply, and that the signaling value at high compliance levels (which is typical for social norms) is rather low. Therefore, although signaling can explain some compliance with social norms, it is unlikely to be the main explanation.

Keywords: social norms, costly signaling theory, signals, Posner, sanctions

Social norms are pervasive features of our lives. They influence almost all aspects of our social interactions, dictating things as disparate as if and how we queue to get on public transport, what counts as a bribe, how we dress, how we greet each other, what demands can be made in the name of family or friendship, what should be done if you are insulted, and how we achieve cooperation in the face of partially conflicting interests. No theory of social life can afford to completely ignore social norms. Many have put forward theories according to which social norm compliance can be explained as signaling behaviour: actors comply with social norms in order to signal that they have particular character traits.[1] The theoretically most well-developed such theory, put forward by Eric Posner,[2] holds that people comply with social norms in order to signal that they are reliable cooperation partners. But I will argue that the case for Posner's theory is overstated.

There are several accounts of how we should define the term "social norm" in more detail. Most such accounts emphasize that social norms are *behavioural regularities*: they concern behaviour, but as the term "social" indicates, it is behaviour that everyone or almost everyone performs, not a particular individual's unique behaviour. For example, it is a social norm that people bring flowers or a bottle of wine when invited for dinner. The literature usually also takes social norms to include perceived *normative requirements*: it is not just

1. See, for example, Ariely, Bracha, and Meier 2009; Hyde 2011; Janssen and Mendys-Kamphorst 2004; Moore 2003.
2. Posner 1998; Posner 2000.

that people tend to bring a bottle of wine when invited for dinner, they think they *ought* to bring a bottle of wine. Social norms concern behaviour that is both wide-spread and widely accepted as normatively correct. However, a social norm can in some special cases also be said to exist even if almost nobody complies with it or think it is normatively correct, as long as almost everyone *believes* that almost everyone complies with it or think it is normatively correct, respectively.[3] More detailed analyses usually also emphasize that social norms are characterized by a set of *expectations* about what others are likely to do, what others expect you to do and think you ought to do.[4] You expect others to bring a bottle of wine when invited to dinner, you expect them to expect you to bring a bottle of wine when you're invited to dinner, and you expect that they think you ought to bring a bottle of wine when invited to dinner. Some more detailed accounts of social norms also point out the importance of the *symbolic meaning* of certain behaviours given particular social norms.[5] The importance of bringing a bottle of wine has little to do with providing the hosts with something to drink, rather, the wine is a gift to show your appreciation of the invitation. Finally, all standard accounts of social norms emphasize that social norms come with a *sanctioning system*: people sanction social norm violators in various ways ranging from glaring at them or telling them off, to negative gossip, social ostracism or perhaps even violence. Such sanctions, or even just the threat of sanctions, are central to maintaining a high level of social norm compliance, and numerous studies have shown that people are willing to impose such sanctions on others.[6] A good definition of social norms needs to address these different characteristics of social norms, and it helps us distinguish between social norms and other types of norms, such as moral norms and legal norms. For example, social norms are often distinguished from *moral* norms by the fact that these expectations about what others do and think are part of what motivates people to comply with social norms, but moral norms are norms one complies with independently of what others do and think.[7] Social norms are distinguished from *legal* norms, in turn, by the fact that the sanctioning system is *informal* in the case of social norms, but for-

3. Suppose our example had concerned behaviour that takes place away from the public gaze: then it might not have been the case that people knew if others in fact performed it. But if almost everyone *believe* that others do so and think they ought to do so, we still have a case of a social norm.
4. See, for example, Bicchieri 2006.
5. For example, see Brennan, Eriksson, Goodin, and Southwood 2013.
6. For example, see Falk, Fehr, and Fischbacher 2005; Fehr and Fischbacher 2004.
7. Of course, in reality, the distinction between social and moral norms is often blurry. One person might comply with a norm regardless of what others do and think while someone else only does so if others also comply and/or think one should comply, and people's moral views tend to be sensitive to the views and actions of others around them, especially over time.

mal in the case of legal ones. If you violate a social norm, you are sanctioned by your peers, without any formal process. But if you break the law, there is a formal process determining what sanctions are to be applied, when and by whom.[8]

But during the last two decades, a different theory of how to explain social norms has emerged in the fields of legal studies and legal philosophy. Eric Posner[9] has suggested that social norms are the aggregate results of people engaging in signaling behaviour. By behaving in a particular way, we signal to other people that we are reliable cooperation partners. For example, he argues, the reason why Americans honour the American flag is that they want to signal that they are reliable cooperation partners. Consequently, they are suspicious of Americans who do not honour the American flag. Many cases seem to fit his theory quite well. Consider for example how teenagers adopt certain forms of dress in order to signal their group membership and their loyalties, thereby informing the other group members that they are reliable cooperation partners.

Posner defines social norms as the aggregated outcome of signaling behaviour. This definition stems from his account of how social norms work. People are looking for reliable cooperation partners. Those people who have a low discount rate for future benefits and costs are more reliable cooperation partners than those who have a high discount rate, because the latter will be easily tempted by short term gains to be made from cheating. Therefore, people look for potential cooperation partners with low discount rates, and have an incentive to signal to others that they themselves have low discount rates. Complying with a social norm often requires that you put your short term benefit aside, something which you are willing to do only if you have a low discount rate and thus prioritize long term benefits. As a result, complying with a social norm can signal your discount rate. Posner therefore argues that people comply with social norms in order to signal their discount rates, not because they have internalized the social norm, or out of habit, a wish to conform, or fear of sanctions. He thus takes social norms as a phenomenon to be *explained* by people's actions to signal their discount rates. He also suggests an account of how social norms emerge. Norm entrepreneurs suggest a behaviour that might work as a signal. If people latch on and begin to understand the behaviour in question as signaling behaviour, a social norm develops. But not all suggestions are taken up, and some signaling behaviour loses its communicative function, in which case people no longer engage in that behaviour. When that happens, norm entrepreneurs suggest a new possible signaling behaviour. One can of course accept Posner's definition of what social norms are and his explanation of why people comply with them, without accepting his particu-

8. See, for example, Brennan et al. 2013.
9. Posner 1998, 2000.

lar theory of how they emerge. Similarly, one can accept the theory of norm entrepreneurs without accepting the theory that social norms are the result of signaling behaviour.

The signaling theory has been applied to social norms by a wide range of theorists from different fields, with or without explicit mention of Posner.[10] Nevertheless, I will argue that Posner has overstated his case: some compliance with social norms can be explained as signaling behaviour, but signaling cannot be the main explanation. Posner himself does not argue that all social norms fit his theory perfectly, and therefore have no problems acknowledging that there might be social norms that do not. However, I will argue that there is a more fundamental problem: social norms are, as a result of their key characteristics, inefficient as signaling devices. Social norms are characterized by high compliance levels, but at high compliance levels, the signaling value of compliance is low. The existence of a sanctioning system against norm violators makes the signaling value even lower because it makes the signal ambiguous. When, on the other hand, very few people comply, the signaling value of complying might be high, but we no longer have a social norm, according to most definitions of social norms. Further, the signaling value even in such a case might be smaller than expected because the signal is ambiguous. Therefore, the main explanation of social norm compliance has to be found elsewhere.

Costly Signaling Theory

Posner's theory is based on *costly signaling theory*, a theory that stems from the field of evolutionary biology. Finding cooperation partners is an important task for both humans and other animals. But the task can be difficult due to two problems. First, you want to be able to identify those who are reliable as partners, so that you don't get cheated. Second, you also want them to choose you among all the other people they could choose to cooperate with. Posner's theory is about the second problem: that of making potential cooperation partners choose you instead of someone else. Suppose people only want to cooperate with those who have a particular trait *T* (for example, the trait of not cheating cooperation partners). You have trait *T* and would thus like to tell them. But why would they believe you? Since everyone wants to get chosen, everyone will say they have trait *T*, regardless of whether they do in fact do so.

10. See, for example, the already mentioned Ariely et al. 2009; Hyde 2011; Janssen and Mendys-Kamphorst 2004; but also, as examples of the wide-spread use of signaling theories in economics, Avery and Levin 2010; Kreps, Milgrom, Roberts, and Wilson 1982; Lewis 2010; and the classic text, Spence 1973.

This is where costly signals come in. Suppose there is a behaviour B that is *costly* for everyone, but *less so* for those who have trait *T*. The people who perform behaviour B are therefore more likely to have trait *T* than people who don't perform behaviour B. A common example in the literature is that of the impressive plumage of male peacocks. The long and beautiful tail feathers seem to serve no function for the peacocks: in fact, they are costly to develop and maintain. But it's the cost that is the key to the explanation of why male peacocks invest resources into developing and maintaining these feathers: it is costly for all peacocks to do so, but much more so for unhealthy, weak peacocks than for healthy, strong ones, because the former can't manage the extra strain on their resources. Investing resources into developing and maintaining such feathers can therefore serve as a signal to females that the male is healthy and strong and thus an attractive mating partner. Notice however that it is crucial that weaker peacocks cannot afford the investment: if all peacocks could afford to develop impressive tail feathers, females would no longer pay attention to the males' plumage, because it would no longer be a credible signal of health and strength. It is the fact that the behaviour is costly, but differentially so, that makes the signal credible. *Receivers* of signals (those for whom the message is intended) generally only pay attention to credible signals, that is, signals that have the kind of cost structure that makes it likely that the *senders* of the signals have the traits they claim to have.

The literature on signaling occasionally distinguishes between those signals that are constitutive of that which is being signaled (the fact that I breathe is a signal that I am alive, because breathing is part of being alive), and those signals that are signs of something *else* (having long and beautiful tail feathers is not part of what it means to be strong and healthy, but it can be a signal that one is strong and healthy because only strong and healthy peacocks can afford long and beautiful tail feathers). Although important in some contexts, the distinction does not matter for the argument to come, and I will therefore not distinguish between the two kinds of signals in what follows.

Costly Signaling Theory Applied to Social Norms

The theory of costly signaling is very influential in evolutionary biology and in parts of the social sciences. For example, it has been suggested that signaling plays an important role in the evolution of human altruism.[11] Posner's suggestion is that it can also successfully be applied to the phenomenon of social norms.

Some things we do generate quite immediate benefits, and others only pay off in the long run. Most of us give less weight to future benefits than we do

11. Fehrler and Przepiorka 2013; Gintis, Smith, and Bowles 2001.

to more immediate benefits. However, how much less differs from person to person: some have a high discount rate for future benefits (that is, they give much less weight to future benefits than to immediate ones), and others have a lower discount rate (they give only a little less weight to future benefits than they do to immediate ones). Posner argues that people with a high discount rate are unreliable as cooperation partners, because the benefits they can get from maintaining a long-term cooperative relationship are in the future and therefore carry much less weight for them than the more immediate benefits they can get from cheating their partner. People with low discount rates, in contrast, will give a greater weight to the future benefits realizable through a long-term cooperative relationship than people with high discount rates do, and will therefore withstand temptations to cheat.

The question is then how people with low discount rates can inform potential cooperation partners about their character. Social norm compliance, Posner argues, is a way of doing that, because social norm compliance is costly (at least generally). The social norm against littering requires us to pick up our garbage after our picnic in the park and put it in a bin, even though leaving it on the ground is more convenient. For those who give weight to long-term benefits and costs, putting the garbage in a bin is still usually worth it: the long term benefit of a clean park and good relationships with the neighbors outweigh the short-term cost. But for those who focus on the short-term benefits and costs, it is not worth it: the short-term inconvenience outweighs the long-term benefits. Thus, because social norm compliance is costly in the short run, people can signal their low discount rates by complying with such norms.

Note, however, that the analogy between social norm compliance and the male peacocks' impressive tail feathers is limited. The signal sent by the male peacock's plumage is not binary. It is thus not the case that peacocks either develop an impressive tail, or no tail at all. Rather, the male peacocks develop as impressive a tail as they can manage. Some will manage more, others less, but it is a matter of degree rather than of all-or-nothing. Social norm compliance, in contrast, is usually binary: either you comply with the norm or you don't. In some cases, it's possible to comply to some extent but not completely, as when you pick up most of your garbage in the park but not every little scrap of paper. But in most cases, you either do or you don't. For example, you can't pick up after your dog to some extent but not fully: every time you pick up, you comply with the norm, every time you fail to do so, you violate the norm. Every time, you either comply completely or violate the norm completely. It doesn't make sense to pick up half of what your dog just left on the ground. That said, there can also be cases in which social norm compliance does come in degrees: it is a social norm that we share a piece of cake equally, but if I

give myself a slightly bigger piece than I give you, I have violated the norm to a lesser degree than if I give myself almost all of the cake.

So far, we've discussed Posner's theory of why people comply with social norms. However, Posner also argues that (at least some) social norms are nothing but the aggregate results of such signaling behaviour.[12] This is a more radical position than the position that people comply with social norms in order to signal, and therefore I am going to concentrate on the latter, less controversial position. It is also this position that is mainly shared by other proponents of theories according to which agents comply with social (or moral) norms in order to signal their character. However, these theories differ when it comes to what it is that is being signaled and the suggestions are quite varied. For example, in some, agents are assumed to signal their sense of social responsibility, in others, their general trustworthiness, a desire to democratize one's country[13] or the trait of a country being a good trading partner.[14] Either way, the basic idea is the same: people perform a costly behaviour B in order to signal that they are people of a particular type.

It is also worth noting that although Posner's theory takes signaling to explain both why social norms emerge and why they persist, you could deny that social norms emerge because of signaling, but still accept that they persist because of signaling. Doing this will however not change anything about the arguments that follow.

These accounts of social norms usually treat social norm compliance as if the signaling value of such compliance is independent of the proportion of others who comply. However, it is not. Rather, the signaling value of social norm compliance is frequency dependent in ways that make the signal ambiguous.[15]

The Case of a High Compliance Level

Social norms are usually complied with by a majority, indeed, often by almost everyone. Wide-spread compliance is part of what makes them social norms, and many even define social norms as behavioural regularities. They might not be complied with by the majority of the population, but then many norms are norms for subgroups, and compliance with these norms is a tool for members of these subgroups to signal group membership, or even partly constitutive of

12. Posner 2000.
13. Hyde 2011.
14. Moore 2003.
15. Madison and McAdams have also made the point that social norm compliance is an ambiguous signal. But neither Madison nor McAdams discuss the kind of ambiguity which is dependent on the proportion of other people who also comply, and, as I will argue, the signalling value of social norm compliance is fundamentally frequency dependent in this way. (Madison 2001; McAdams 2001).

group membership. What makes these varieties of dress codes social norms is that they are widely adhered to *in the subgroup in question*.[16]

But such a high level of compliance can mean one of two things, both of which are problematic:

a) Almost everyone has trait *T*, or
b) Even those without trait *T* comply with the social norm.

If almost everyone has trait *T*, the signal is credible: those who comply with the social norm in question—the senders of the signal—do have trait *T*. But if almost everyone has trait *T*, then potential cooperation partners—the receivers of the signal—would not take much of a risk if they just picked a partner randomly: they would be likely to pick someone with trait *T* anyway. If there is some cost involved in detecting social norm compliance, such as paying attention and remembering who complied and who didn't, they are unlikely to find it worth the cost to focus on detecting norm compliance. But that means that the senders of the signal lose their incentive to comply with the social norm: there is not much point sending a signal that nobody pays attention to. In fact, signaling that you have trait *T* is most valuable when few others have trait *T*, and becomes less valuable the greater the proportion of people who have trait *T* is.

16. Some signs of group membership are not so much about signaling that one belongs to the group, as acts of commitment to that group. I am grateful to an anonymous reviewer for pointing this out. However, acts of commitment can be understood in different ways, First, as acts that cause the agent to belong to the group, as when you cause yourself to belong to a group by signing a contract that states that you belong. Second, as an unfakeable sign that one belongs to the group (like speed, strength and/or colourful and costly plumage are unfakeable signs that an animal is healthy). Skin colour or other hard-to-fake physical characteristics can be characteristics on the basis of which humans are classified into groups in a similar way. Finally, acts of commitment can be understood as a signal that the agent will never betray the group because the agent has cut him- or herself off from other groups and so cannot hope to cooperate with members of these other groups. The latter is a form of costly signal. It is not quite appropriate to speak of cases in which we belong to groups in virtue of unfakeable characteristics like skin colour as cases involving social norms, because social norms require that people can choose not to behave in the way the norm specifies. But if we perform an act of commitment by causing ourselves to have the particular, relevant trait that makes us members of a relevant group, then for members of that group, that act might be a social norm, because acts of commitment might be required. If commitment requires a particular action, then performing that action would have a high compliance rate among group members, and it would be considered a social norm for group members to perform that particular act. But if group members can commit in a variety of different ways, then different group members might commit by performing different sorts of actions, neither which will be required by a social norm. Instead, the social norm will be that one commits to the group in some way or another. However, a group might not require that members perform any kind of commitment act. There might still be some actions that are not performed by most group members, but which are still looked favourably upon as acts that honour the group. But these will then not be social norms.

If, on the other hand, everyone complies *regardless* of whether they have trait *T*, the signal is unreliable. If both those with trait *T* and those without trait *T* comply with the social norm in question, those looking for a cooperation partner have no reason to invest any energy or resources in detecting social norm compliance, because social norm compliance is not a reliable sign of having trait *T*. And if nobody believes you when you signal that you have trait *T*, you have no incentive to signal by complying with the social norm.

The outcome in a case in which everyone complies with social norms, regardless of whether they possess trait *T*, is called an *active pooling equilibrium*. It occurs when the cost of complying with social norms is insufficient for generating a difference in behaviour between the people with trait *T* and the people without trait *T*: social norm compliance is simply not costly enough for people without trait *T* to be tempted to violate norms.[17]

In one paper, Posner briefly discusses such pooling equilibria (Posner 1998). When they occur, Posner argues, people will cease to comply with the social norm and norm entrepreneurs will suggest some new social norm that can function as a signal. This undoubtedly sometimes happens, although, as McAdams (2001) have pointed out, norm entrepreneurs typically do *not* tell people to "Do X, because bad types won't," but rather, to "Do X—everyone does!" However, Posner (also) fails to recognize that pooling equilibria are not just an occasional outcome of signaling through social norm compliance, but the *standard* outcome. The reason is the sanctioning system that characterizes social norms. What sanctions against social norm violators do, is to create short-term costs for violating social norms. This makes social norm compliance seem like a good choice even for people who lack trait *T*. The effect is that even those without the relevant trait *T* comply with social norms. A central characteristic of social norms thus drives people to behave in a way that creates a pooling equilibrium.[18] Of course, some of those who comply because

17. Pooling equilibria can arise with regards to tail feathers too. Most cost structures will of course generate a distribution of tail feather length and brightness that ranges from the very impressive to the not so impressive. But if it is cheap enough to develop an impressive tail, then both strong, healthy and weak, sick peacocks will develop one.

18. If the behaviour in question is an unfakeable sign that one has trait T, then sanctions will not have this effect, and will in general be redundant since people without trait T cannot behave as if they do. In some such cases, there will then be no social norm that specifies that one behaves as if one has trait T, because people who do not have trait T cannot behave that way. But behaving as if one has trait T could of course still be a choice for people who do have trait T, that is, an agent with trait T might be able to choose whether to show this to others by behaving in the specified way or not. There could then be a social norm among people with trait T that one performs the behaviour that is associated with having T. For example, it is sometimes argued that there is a social norm among members of the upper class that one complies with complicated etiquette rules to demonstrate that one was indeed born into this class, as these rules are difficult to learn if one has not been brought up in

of the threat of sanctioning might indeed have the relevant trait *T*: They could have preferred to violate the social norm not because of a high discount rate, but because their concern for long-term benefits of social interaction was overridden in that particular instance by other concerns. Those people comply with the social norm in that particular instance because of the threat of sanctions, but still have the trait *T*.[19] Nevertheless, sanctions still make compliance an ambiguous signal.

Posner does in fact notice something that should alert him to this problem: he argues that the social norm for Americans of honouring the American flag loses its significance as a signal when there is a law that forbids dishonouring the American flag.[20] He therefore argues that policy makers should be wary of using laws to enforce compliance with social norms that serve as signals, because such laws might inadvertently undermine the social norm they were supposed to strengthen. But he doesn't notice that social norms come with their own equivalent of such a law, and does so more or less per definition. And that, of course, generally makes social norm compliance a rather inefficient way of signaling.

In fact, if compliance with social norms is a signal, the signal would be undermined by the sanctions against norm violators, because the threat of sanctions makes it difficult to know why people comply with norms. But instead, sanctions against norm violators tend to strengthen norms. They do so in two ways: the emphasis on an obligation to behave in a particular pro-social way has an independent effect on people's preferences, but it also affects people's

---

an upper class environment. By requiring members of the upper class to demonstrate their group membership by complying with these etiquette rules, the upper class is sustained as a group distinct from others in society. In a case like this, we might indeed get a high compliance rate among members of the upper class, maintained by sanctions against members who for some reason are tempted to not comply with the etiquette rules, but where the sanctions are not making people who lack the relevant trait behave as if they do possess it. However, most cases of social norms do not concern cases in which the members of a group are required to perform a certain behaviour that cannot reasonably be performed by non-members as a way of keeping the group distinct from others. If nothing else, staying within Posner's rationalistic framework requires that explanations are based on what it is in people's individual interest to do, not on references to the persistence of groups per se. Notice also that not all unfakeable signs of group membership will be behaviours required by some social norm. Speaking a language fluently because you were born into an ethnic community does not necessarily mean that you will be sanctioned for choosing to express yourself in some other language instead.

19. I am grateful to an anonymous reviewer for this point.
20. The results of other research indicate that he is right to be concerned: for example, Griskevicius et al. show that if so-called *green* products become cheaper than less environmental products, consumption of them often goes down. The reason is that consumption of the green products could be interpreted as the result of financial constraints. (Griskevicius, Tybur, and Van Den Bergh 2010.)

beliefs about what most others are going to do. That means that conditional co-operators can feel assured that others will cooperate, and therefore they cooperate too. [21]

In his book on social norms, Posner discusses sanctioning, but without recognizing the fundamental problem that sanctioning constitutes for the theory of social norm compliance as signaling (Posner 2000). Instead, he focuses on how to explain sanctioning itself. If we comply with social norms to send a signal about our low discount rates and thereby high reliability as cooperation partners, why would we bother punishing those who choose not to send such a signal? Posner's argument is that sanctioning those who violate a norm is itself a form of signaling behaviour, because sanctioning is costly (at least in most forms).[22] Experiments show that third-party sanctioning of norm violators is indeed at least partly about signaling. For example, Jordan et al. has shown that third-party sanctioners are more trustworthy than others, and are also trusted more. But sanctioning is not an optimal form of signaling. If a person X behaves badly to person Y by violating a social norm, and a third party—person Z—observes this, most people who happen to be in the role of person Z will prefer to help person Y directly in some way, rather than punishing person X. Jordan et al. has shown that when third-party observers have the opportunity to help victims of social norm violations directly, their sanctioning behaviour against the social norm violators decreases and their helping behaviour towards the victims increases. When that happens, the signaling value of punishing norm violators decreases: the trustworthy third-party sanctioners tend to prefer to help people directly rather than to punish norm violators, so when direct helping behaviour is available as an alternative, those who still focus on sanctioning norm violators instead of helping directly tend to be the less trustworthy individuals. Observers are aware of this: when direct helping behaviour is an available alternative, people tend to perceive those who nevertheless focus on sanctioning norm violators as less trustworthy.[23]

However, note that in real life, it is rare that our only option for engaging in a pro-social way with others is by punishing norm violators. Usually we have the option of engaging in direct helping behaviour or being cooperative. Therefore, the signaling value of punishing norm violators is likely to be weak in most real-life settings. Just how weak the signal will be will depend on the context. But it seems implausible to think that in most contexts where norms are complied with, they are sustained by the sanctioning threats of people with very few other pro-social behaviour options than punishing norm-violators.

21. Bowles and Hwang 2008.
22. Although, see Brennan and Pettit 2004.
23. Jordan, Hoffman, Bloom, and Rand 2016.

But more importantly, if the reason why people comply with social norms is because they want to signal their low discount rate, then sanctioning would not be needed to sustain such norms. People will engage in signaling whenever the expected benefit of sending a signal is greater than the expected cost of not doing so. By considering sanctioning to be irrelevant for people's decisions to comply with social norms, Posner creates a theory that is quite different from most theories about social norms, since these ascribe a crucial role to sanctioning (and/or to internalization, but Posner does not give a significant role to internalization either).[24]

Notice also that if Posner is right, those who possess the relevant trait *T* would be faced with a dilemma. If they sanction non-compliance, they undermine the signal sent by compliance with the norm, so their own compliance with the norm becomes less valuable. And if compliance with the norm no longer sends the desired signal, sanctioning those who violate that norm will not send that signal either. If, on the other hand, those with the trait *T* choose *not* to sanction norm violators, they avoid undermining the signaling value of norm compliance, but instead they cast doubt on their commitment to the norm.

Further, the risk of being sanctioned for a norm violation is itself frequency-dependent. Sanctioning social norm violators is voluntary and usually costly. Some people enjoy self-righteous criticism of others, but most people find it unpleasant to have to tell people off. Further, sanctioning can involve greater costs than unpleasantness: you risk ruining your relationship with the people you sanction, and they might even retaliate. Therefore, sanctioners are normally in short supply. But when the compliance level is high, there are more potential sanctioners (in most cases, only people who comply with the norm themselves are willing to sanction violations of it). When the compliance level is high, there are also fewer instances of sanctioning required, so for any given number of available sanctioners it is more likely that those available will be sufficient. And finally, because there are fewer norm violations, there is normally a greater risk that any given one will be noticed. Therefore, the threat of getting sanctioned for violating a norm generally increases as the level of compliance with the norm increases. As a result the threat of sanctions is particularly high at high compliance levels, and in such cases it is therefore also particularly ambiguous whether people comply with the norm because of the risk of sanctions, or because they possess trait *T*.

A final reason why the signaling value of social norm compliance is frequency dependent is that many social norms are characterized by a particular form of perceived normative requirement: this is the notion of "doing one's fair share." For example, if others work to keep the environment clean by not littering, you should "do your fair share" to keep the environment clean by also

24. Madison 2001.

abstaining from littering. But this normative requirement of "doing your fair share" only exists if sufficiently many others do their part. Otherwise, there is no "fair share" to be done. As many experiments have shown, people do in fact litter much more in environments that are already dirty than they do in pristine environments.[25] Indeed, people tend to feel much more guilty for not cooperating if others are cooperating, than if others don't.[26] They also tend to take moral cues from their peers, so if many people engage in a behaviour and speak well of it, others are much more likely to come to share the expressed moral view.[27]

The Case of a Low Compliance Level

If almost nobody has trait *T*, signaling that you have *T* can be very beneficial. Others pay attention, because reliable cooperation partners are rare, and therefore, people have an incentive to pay the cost to signal. It might thus seem that Posner's theory would be a more promising explanation for social norm compliance in cases in which the level of compliance with the norm in question is low. However, Posner's theory is supposed to explain social norm compliance, and it is an essential characteristic of social norms that there is wide-spread compliance, or at least the perception of wide-spread compliance. If a behavioural rule does not generate a reasonably high level of compliance, then it is not a social norm.[28] For example, it is considered nice to pick up after your

25. Cialdini, Kallgren, and Reno 1990.
26. Miettinen and Suetens 2008: 952.
27. Cialdini 1989.

    Although Posner's argument concerns cooperation norms, we could apply it to coordination norms as well. The reason why most arguments about signaling tends to focus on cooperation is probably because signaling only is relevant if there is a reason for people to want to be chosen by others, and because most coordination concersn cases in which coordination happens without anyone having to be chosen for anything. But suppose we have a case in which interests are not in conflict but people want to coordinate, and suppose different groups coordinate on different strategies etc. You need to figure out with whom to coordinate. In a small group, you just raise your voice and ask everyone in "your" coordination group (however that is defined) to raise their hand, but in wider society that is not so easy. You thus need a way for members of "your" coordination group to identify themselves. Here, too, there would be no problem of fake signals, because in coordination games, interests are not in conflict, and neither would there be a need for a sanctioning system, because nobody wants to coordinate with any other group than their own, and nobody else wants to force them to do so, because that would distort a successful coordination system. I am grateful to an anonymous reviewer for prompting me to raise the question of coordination norms.
28. Strictly speaking, what is required is that almost everyone believes that almost everyone complies, not that they actually do it. But the kind of behaviours that are most relevant for signaling are behaviours performed in public, and people are therefore unlikely to be mistaken about whether almost everyone performs the behaviour in question.

dog, but it is not a social norm to do so unless almost everyone does this and is expected to do it. That said, even if the behaviour in question is not widespread enough to be a social norm, it can work quite well as a way of signaling.

However, because the notion of "doing one's fair share" depends on whether there is something like a fair share to be done, and because that in turn depends on the perceived proportion of people who comply with the social norm, complying in the case of a widely known low compliance level can also be interpreted in more than one way. Consider the normative requirement to wash your dirty coffee mug instead of putting it in the sink in the staff tea room. Everyone wants the tea room to look nice, but everybody thinks that washing their coffee mugs is inconvenient. If others are actually keeping the tea room clean, you are normatively required not to ruin it; the others are doing their fair share, and so should you. But if the others are leaving their dirty coffee mugs all over the kitchen bench, then they are not doing their "fair share," and it would be ridiculous to hold you normatively required to carefully wash your mug every time, since there is no cleanliness to preserve and no "fair share" to be done. But in that case, those who wash their coffee mugs when almost nobody else does so are not doing so in order to comply with the normative requirement of "doing their fair share." They might do it in order to *create* a social norm by pointing out the potential for cooperation. They might be acting out of habit (presumably their own kitchen is nice and clean). Or they might be signaling to others that they are exceptionally tidy and/or socially responsible people. But they are certainly not "doing their fair share," because the normative requirement of "doing one's fair share" does not apply. Many social norms include an element of "doing one's fair share": the social norm not to litter, the social norm to pick up after your dog, the social norm to refrain from tax evasion[29], the norm to be polite in traffic, the norm not to push ahead in the queue, the norm to pay when it's your turn to buy everyone a beer, and so on.

Indeed, washing your coffee mug when few others do can be interpreted in ways you don't want. It can be interpreted as showing that you're easily exploitable (a "sucker"), or that you wish to "seem better" than others, or that you think the others are slobs. Indeed, it is quite likely that the more salient your compliance is—that is, the more demonstratively you wash your mug in front of others and make sure they notice you—the more likely it is that you will be considered self-righteous and critical rather than responsible and trustworthy.

29. Refraining from tax evasion is of course also a legal norm, but legal and social norms can come apart. There might be a law against tax evasion, but a social norm that one tries to cheat the system as best as one can.

Even More Ambiguity

There is (at least) one more way in which social norm compliance can be explained, which further increases the ambiguity of norm compliance as a signal: by complying with social norms, I signal that I am the kind of person who complies with social norms. That does make me reliable: I can be relied upon to comply with whatever the social norms are.[30]

Indeed, one might argue that if what matters is my reliability, it is not very important what reasons I had for complying. What matters is that I can be relied upon to comply with social norms. [31] But being reliable in the sense of reliably complying with social norms has somewhat different implications than being reliable in the sense of having a low discount rate for future costs and benefits. The latter says something about how you will behave even in situations in which there are no clear social norms, and in some cases, you might even violate a social norm for the sake of a long-term, cooperative relationship.[32] The fact that social norm compliance might just as well signal that one is the kind of person who complies with social norms thus matters, from the point of view of the receiver, who is trying to interpret the message. And it is impossible to tell from the mere compliance with social norms whether the person who complies has a low discount rate for future benefits, or whether s/he just is the kind of person who generally complies with social norms.

Possible Objections

Poisonous frogs are usually brightly coloured. Their colours signal that they are poisonous, and predators interpret the signals accordingly and avoid attacking the frogs. But the signal is not completely reliable, since there are non-poisonous frogs that mimic the bright colours of the poisonous ones. Yet, that does not ruin the signal, or make the explanation of the bright colours of poisonous frogs less valid. Therefore, one might argue that social norm compliance by people with a high discount rate would not undermine social norm compliance as a signal of a low discount rate.

30. However, in the peacock case, there is a clear distinction between what the beautiful feathers signaled (health, strength) and the means through which that message was conveyed to females (beautiful, costly feathers). In the case of social norm compliance, that distinction threatens to collapse. Part of what it means to be reliable, is to comply with social norms. Therefore it's not clear that complying with social norms is a signal of reliability rather than partly constitutive of reliability. I owe this point to Garrett Cullity.
31. I owe this point to Garrett Cullity.
32. Taking "reliability" to mean, as Posner does, "having a low discount rate for future costs and benefits," you get a clearer distinction between what is being signaled and the means through which the message is conveyed than if you take "reliability" to mean "complying with social norms."

But notice what happens in the case of the frogs: the value of the signal does go down when non-poisonous frogs mimic the colours of the poisonous ones. After a while, when the mimicking frogs are sufficiently many in proportion to the poisonous ones, predators will notice that the risk of eating a brightly coloured frog has decreased, and they will begin to eat more such frogs. Since it no longer is as valuable to be brightly coloured, mimicking frogs have less incentive to mimic the bright colours (and the poisonous frogs have less incentives to produce such colours too). In fact, being highly visual is a real problem if predators begin to hunt frogs with bright colours. Eventually, there is an equilibrium between the proportions of poisonous frogs and of mimicking frogs.

The same sort of dynamic might be said to apply to social norms. When complying with social norms has a sufficiently large signaling value, even those with high discount rates will comply with norms: even highly discounted, the future benefits are large enough to outweigh the short-term costs of compliance. This will lower the signal value, because people will notice that both those with low and those with high discount rates comply with social norms. In the end, an equilibrium might get established in which some proportion of people with low discount rates do not comply with norms, and some proportion of people with high discount rates do. Such an equilibrium does of course not by itself show that signaling is not a good explanation of social norm compliance. But since the equilibrium is unlikely to be at the level of near-full compliance, it does suggest that near-full compliance is difficult to explain in terms of signaling. Since social norms are generally characterized by near-full compliance (or perceived near-full compliance)[33], this is bad news for the theory of social norm compliance as signaling.[34]

Another possible objection concerns the signaling value of non-compliance. I have argued that at high compliance levels, the signal sent by complying with the norm in question is ambiguous. But the negation of something ambiguous might not itself be ambiguous. It could be the case that even if we cannot interpret norm compliance unambiguously, failure to comply with the norm might be a clear signal. As Axelrod puts it,

33. As stated earlier, most accounts of social norms stress that they are defined partly by the fact that they are behavioural regularities, or at least perceived behavioural regularities.
34. Of course, the exact level of compliance at equilibrium depends on the cost of complying. In the case of frogs, the cost is the higher visibility to predators that bright colours result in. In the case of norms, the cost will vary depending on the individual norms and contexts. In some situation, compliance with social norms might be costly indeed, in other situations, the cost might be nothing more than the slight effort of putting your garbage in the bin rather than drop it on the ground.

> An important, and often dominant, reason to respect a norm is that violating it would provide a signal about the type of person you are.[35]

However, in the case of norm compliance, failure to comply is ambiguous too. It can indicate that the agent was not reliable, that s/he as not afraid of sanctions, that s/he stands up against peer pressure, or that s/he has sufficiently high social status to not have to worry about getting sanctioned. Further, non-compliance with a social norm of one group can constitute compliance with the social norm of another group.[36] Teenagers who dress only in black and colour their hair green and people who tattoo their faces or their hands comply with the social norms of a subgroup, even as they fail to comply with the social norms of the larger population. Indeed, the main point about many sub-group norms is usually to show loyalty by creating a look that distances the members from the main population and thereby decreases the opportunities for cooperation with the main population.

A third possible objection is that we might develop Posner's general idea in a way that makes it less vulnerable to the criticism that I've outlined here. Druzin argues that Posner is wrong to treat a person's social norm compliance as a series of individual decisions about when to signal. It is not the case, Druzin argues, that a person P decides to comply with a norm N in order to signal her low discount rate. Rather, people with low discount rates develop a general tendency to internalize the normative requirements of norms, and to comply with norms in general. Indeed, the tendency to internalize normative requirements makes people more efficient signalers, since they signal their low discount rates instinctively instead of having to plan it each time.[37]

But Druzin's argument does not really solve the main problem for Posner's theory. It is still the case that if compliance levels are very high, the signaling value of compliance is undermined. This problem does not change even if the reason people comply is because they have internalized normative requirements rather than because they strategically plan how to signal their discount rate.

Further, it seems very reasonable to assume that we are the kind of creature we are because it has been evolutionary beneficial to be that kind of creature. And the reason why it has been evolutionary beneficial to be the kind of creature who internalizes socially transmitted normative requirements about behaviour *could* very well be the need for people to determine who is a reliable cooperation partner. But it might also be the case that we are the kind of creature we are because learning social rules for behaviour and internalizing these

35. Axelrod 1997: 62.
36. For example, see Brennan et al. 2013 and Druzin 2013. I am grateful to an anonymous reviewer for suggesting that I include this point.
37. Druzin 2011.

as normative requirements makes us reliable cooperation partners and enables us to function efficiently in a group. If that is the case, social norm compliance is not about signaling per se.

One could also interpret a person's decision to comply with a social norm as a signal not of their discount factor for future interactions, but rather just as a signal of the probability that they will be around for future interactions.[38] Profitable long-tem relationships require that the agent is a reliable cooperation partner, because if the agent cheats someone, they will not want to interact with him or her again. But how does the agent signal their intention of staying around for a long time and therefore their commitment to long-term interactions? By complying with social norms. Complying with social norms would not be worth it to the agent if they were only here for a short time, so compliance signals that they are going to be around for the foreseeable future. But my argument against signaling as the main explanation would work against this interpretation too. Suppose almost everybody complies with social norms. If everybody is around for the foreseeable future, and observers care about finding people who are around for the foreseeable future rather than about norm compliance per se, then there is no point for them to care about who complies with social norms and who doesn't. If, on the other hand, almost everybody complies but a significant proportion of those who do are *not* going to stay around for long-term interactions, observers will soon figure out that compliance is not a credible signal of long-term commitment, and will therefore stop paying attention to social norm compliance.

Finally, yet another possible objection concerns the definition of a signal. You can of course behave in a way intended to communicate a message to someone, but fail to achieve the desired outcome because the intended receiver of the message did not understand, or did not pay attention. This is indeed what happens when you attempt to explain the details of something to someone who lacks the necessary background knowledge and therefore doesn't understand your explanations. Even though the goal wasn't achieved we still explain your behaviour in terms of your wish to communicate a particular message. That might make us define the signal as the message you wanted to communicate, and explain your behaviour as signaling, even though your goal of getting the message across to the intended receiver was not achieved. If we define signaling behaviour this way, you might argue that one can still understand a person's social norm compliance as signaling behaviour even in a situation in which almost everyone else also complies. Everybody is trying to tell others that they have trait T, but nobody listens because both those with and those without trait T are claiming to have the trait in question. However, we would not normally characterize as a signal behaviour that was not socially understood to normally

38. I owe this point to an anonymous reviewer.

communicate a message successfully. You might fail to communicate because the particular person you wanted to communicate with did not understand, or did not pay attention. But if there was a general lack of shared understanding or attention, what you did cannot be explained as signaling. Further, people engage with social norms on a regular basis. Anyone can try to send a signal that there is a bear nearby by walking in a small circle while hooting. But after the message fails to get understood the first time, you are unlikely to keep trying to get the message across this way.

Conclusion

It is probably true that some social norm compliance is motivated by a wish to signal something about yourself, be it having a low discount rate or something else. But it is implausible that signaling can be the whole or even the main explanation for social norm compliance, given that social norms are characterized partly by in general high levels of compliance and by the fact that sanctioning plays an important role in sustaining such high levels of compliance. Although Posner does admit that he doesn't intend his theory of social norm compliance as signaling behaviour to be understood as a theory about all social norms, his work lacks an appreciation of just how badly standard cases of social norms fit his theory.

Flinders University and University of Gothenburg
lina.eriksson@gu.se

Acknowledgments

I am grateful to three anonymous reviewers for helpful comments.

References

Ariely, D., A. Bracha, and S. Meier. 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review* 99(1): 544–55. https://doi.org/10.1257/aer.99.1.544

Avery, C., and J. Levin. 2010. "Early Admission at Selective Colleges." *American Economic Review* 100: 2125–56. https://doi.org/10.1257/aer.100.5.2125

Axelrod, R. 1997. *The Complexity of Cooperation: Agent-Based models of Competition and Collaboration*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400822300

Bicchieri, C. 2006. *The Grammar of Society*. Cambridge: Cambridge University Press.

Bowles, S., and S. Hwang. 2008. "Social Preferences and Public Economics: Mechanism Design when Social Preferences Depend on Incentives." *Journal of Public Economics* 92(8–9): 1811–20. https://doi.org/10.1016/j.jpubeco.2008.03.006

Brennan, G., L. Eriksson, R. Goodin, and N. Southwood. 2013. *The Explanation of Norms*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199654680.003.0002

Brennan, G., and P. Pettit. 2004. *The Economy of Esteem*. Oxford: Oxford University Press. https://doi.org/10.1093/0199246483.001.0001

Cialdini, R. 1989. "Social Motivations to Comply: Norms, Values and Principles." In *Taxpayer Compliance*, vol. 2, ed. J. Roth and J. Scholz, 213–14. Philadelphia: University of Pennsylvania Press. https://doi.org/10.9783/9781512806281-008

Cialdini, R., C. Kallgren, and R. Reno. 1990. "A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior." *Advances in Experimental Social Psychology* 24: 201–34. https://doi.org/10.1016/S0065-2601(08)60330-5

Druzin, B. 2011. "Law, Selfishness, and Signals: An Expansion of Posner's Signaling Theory of Social Norms." *Canadian Journal of Law and Jurisprudence* 24(1): 5–53. https://doi.org/10.1017/S0841820900005051

Druzin, B. H. 2013. "Eating Peas with One's Fingers: A Semiotic Approach to Law and Social Norms." *International Journal for the Semiotics of Law* 26: 257–74. https://doi.org/10.1007/s11196-012-9271-z

Falk, A., E. Fehr, and U. Fischbacher. 2005. "Driving Forces behind Informal Sanctions." *Econometrica* 73(6): 2017–30. https://doi.org/10.1111/j.1468-0262.2005.00644.x

Fehr, E., and U. Fischbacher. 2004. "Third-Party Punishment and Social Norms." *Evolution and Human Behavior* 25(2): 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4

Fehrler, S., and W. Przepiorka. 2013. "Charitable Giving as a Signal of Trustworthiness: Distentangling the Signaling Benefits of Altruistic Acts." *Evolution and Human Behavior* 34(2): 139–45. https://doi.org/10.1016/j.evolhumbehav.2012.11.005

Gintis, H., E. A. Smith, and S. Bowles. 2001. "Costly Signaling and Cooperation." *Journal of Theoretical Biology* 213: 103–19. https://doi.org/10.1006/jtbi.2001.2406

Griskevicius, V., J. M. Tybur, and B. Van Den Bergh. 2010. "Going Green to Be Seen: Status, Reputation, and Conspicuous Conservation." *Journal of Personality and Social Psychology* 98(3): 392–404. https://doi.org/10.1037/a0017346

Hyde, S. 2011. *The Pseudo-Democrat's Dilemma: Why Election Monitoring became an International Norm*. Ithaca, NY: Cornell University Press. https://doi.org/10.7591/cornell/9780801449666.001.0001

Janssen, M., and E. Mendys-Kamphorst. 2004. "The Price of a Price: On the Crowding Out and In of Social Norms." *Journal of Economic Behavior and Organization* 55(3): 377–95. https://doi.org/10.1016/j.jebo.2002.11.004

Jordan, J. J., M. Hoffman, P. Bloom, and D. G. Rand. 2016. "Third-Party Punishment as a Costly Signal of Trustworthiness." *Nature* 530: 473–76. https://doi.org/10.1038/nature16981

Kreps, D., P. Milgrom, J. Roberts, and R. Wilson. 1982. "Rational Cooperation in the FinitelyRepeated Prisoners' Dilemma." *Journal of Economic Theory* 27(2): 245–52. https://doi.org/10.1016/0022-0531(82)90029-1

Lewis, G. 2010. "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors." *American Economic Review* 101(4): 1535–46. https://doi.org/10.1257/aer.101.4.1535

Madison, K. 2001. "Government, Signaling, and Social Norms." *University of Illinois Law Review* 2001: 867–85.

McAdams, R. 2001. "Signaling Discount Rates: Law, Norms and Economic Methodology." *The Yale Law Journal* 110: 625–89. https://doi.org/10.2307/797597

Miettinen, T., and S. Suetens. 2008. "Communication and Guilt in a Prisoner's Dilemma." *Journal of Conflict Resolution* 52: 945–60. https://doi.org/10.2139/ssrn.1092787

Moore, D. M. 2003. "A Signaling Theory of Human Rights Compliance." *Northwestern University Law Review* 97(2): 879–910.

Posner, E. 1998. "Symbols, Signals, and Social Norms in Politics and Law." *Journal of Legal Studies* 27: 765–98. https://doi.org/10.1086/468042

Posner, E. 2000. *Law and Social Norms*. Cambridge, MA: Harvard University Press.

Spence, M. 1973. "Job Market Signaling." *The Quarterly Journal of Economics* 87(3): 355–74. https://doi.org/10.2307/1882010