THE STATE OF NATURE
Michael Smith

1. What Is the State of Nature?

As I understand it, the state of nature is the state of a world without any social rules. Thinking about the state of nature is important because it focuses our attention on the normative reasons for action we have—from here on all talk of reasons for action will be about normative reasons—in circumstances in which there are no social rules, and hence reveals the reasons there are, if such reasons exist, for creating and maintaining such rules.

To be clear, what is of interest is not what the reasons were for creating the social rules that we live under, where it is being assumed that these rules were created by people for what they took to be reasons, as no such assumption is being made. What is of interest is rather the normative significance of social rules, both actual and possible. Thinking about states of nature reveals this to us precisely because it allows us to abstract away from historical questions about how the actual rules came into existence.

For these purposes, social rules can be understood in the way proposed by H.L.A. Hart at the beginning of The Concept of Law (1961). According to Hart, a social rule is a regularity in the behavior of the members of a social group, where that regularity is kept in place by the members of a sub-group who appeal to it in criticizing the behavior of those who fail to conform and affirming the behavior of those who do. The members of the sub-group thus occupy a certain position of power vis-à-vis the group as a whole. Hart puts no constraints on what members of the sub-group can believe their reasons are for going in for such criticism and affirmation, or what their power might consist in, and nor does he require that they all believe that they have the same reasons. The only constraint is that the reasons they believe they have do not themselves presuppose the existence of the social rule whose existence they help explain.

A social rule thus presupposes: (i) the existence of a regularity in the behavior of a group; (ii) members of a sub-group who believe that they have reasons to criticize the behavior of those who fail to conform to that regularity and affirm the behavior of those who do, where the reasons they believe they have do not themselves presuppose the regularity’s existence; and (iii) a causal
relationship between (ii) and (i). Uncontroversial examples of social rules, so understood, are the rules of games and the (changing) rules of fashion. Only slightly more controversial examples include the rules associated with different languages, and the rules associated with different legal jurisdictions. The immediate significance of social rules, so understood, is clear when we think about Hart’s account of the last of these.

If we think of a legal system as constituted by a distinctive set of social rules, as Hart does, then the task of the legal theorist is to explain what the distinctive social rules are that constitute a legal system, and how such rules could emerge from a pre-legal world and be maintained once they have emerged. What sorts of reasons might the members of the sub-group believe they have for going in for criticism and affirmation, and do such reasons exist? And what reasons might the members of the social group who aren’t members of the sub-group believe they have for doing what they do when they begin acting in the way that will become a regularity, and for conforming to it once it becomes a regularity, and do these reasons exist?

If legal theorists can complete this task then they will thereby have explained the normative significance of law in terms of the reasons that there are to establish and conform to the relevant social rules. If they can only complete half the task, because (say) it turns out that there aren’t the reasons that people believe there are, then the law will turn out to be normative in one sense but not in another. In Tristram McPherson’s helpful terms, the law will be formally normative, but not authoritative [McPherson 2011]. This last point is worth dwelling on. If the state of nature is the state of a world that exists prior to the existence of any social rules, then it follows from Hart’s account of what a social rule is that we might be able to tell a story about the emergence of social rules without having told a story about the reasons that there are to establish and conform to them. Indeed, legal positivism exploits just this possibility.

The states of nature we need to think about to explain the normativity of social rules are thus a subset of all of the states of nature. For example, it would be pointless to think about those states of nature in which there are no people. Nor will it be helpful to think about those states of nature in which there are people, but they are especially benighted or apathetic or weak-willed, at least not in the first instance. The states of nature we need to think about are those in which there are people; there are reasons for those people to act in certain ways; the people in the state of nature know what these reasons are; they act on them in a robust way; but there are no
regularities in their behavior that are kept in place by the members of a sub-group who appeal to those reasons in criticizing the behavior of those who deviate from the regularities and affirming the behavior of those who conform.

Restricting our attention to these states of nature, at least in the first instance, allows us to put to one side two very different kinds of question to which we need answers in understanding how and why social rules emerge and are maintained. These are questions about motivation and questions about justification. The questions about motivation are: What motivates those in the sub-group who bring about the regularity in the behavior of the group as a whole to bring about that regularity, and what motivates those in the group as a whole to act in the way that eventually becomes a regularity? The questions about justification are: What reasons do people in the sub-group have to bring about a regularity in the behavior of the group as a whole, and what reasons do those in the larger social group have to act in the way that eventually becomes a regularity? Since people can be motivated to bring into existence social rules that they have no reason to bring into existence (they might be benighted), and they can have reasons to bring social rules into existence but not be motivated to do so (they might be apathetic or weak-willed), these two questions are clearly different.

Initially restricting our attention to those states of nature in which people know what they have reasons to do and are robustly motivated to act on those reasons, allows us to put the difference between questions of justification and motivation to one side. Doing so will clarify whether social rules are only significant because of the possibility of our failing to do what we recognize we have reason to do, or our failing to respond to our recognition of that fact by being motivated accordingly, or whether social rules would still be significant even if we didn’t fail in one of these ways.

2. What Reasons for Action Do People Have in the State of Nature?

Any answer to the question of what reasons people have in the state of nature is bound to be controversial. Let me begin by giving the best answer, and then I will briefly explain why it is the best answer.

States of nature are just one among many circumstances in which people might find themselves with reasons to act, where what makes it true that people have reasons to act in certain ways, no matter what circumstances they find themselves in, is the desirable ways the
world would be if they were to act in those ways in those circumstances. This is what Joseph Raz calls the ‘classical conception’ of reasons for action (Raz 2002). According to the classical conception, all reasons for action are grounded in the desirable states of the worlds that result when actions are performed, where the desirable feature could just be the performance of the act itself. This allows us to rephrase our question. What is the ordering of states of nature from most desirable to least desirable?

The best answer to this question, it seems to me, is perspectival. From the perspective of any particular agent, there are three desirable ways the world could be. We could think of these as all bound up with the capacity to lead a life of one’s own choosing. What’s desirable, from the perspective of any particular agent, is that the world be one in which they lead a life of their own choosing, free of ignorance and compulsion (from here on I will take the ‘free of ignorance and compulsion’ condition as read); that the world be one in which they don’t interfere with anyone’s leading a life of their own choosing, so long as their leading that life doesn’t require their interfering with someone else’s leading a life of their choosing (from here on I will mostly take this condition as read); and that the world be one in which everyone has the wherewithal to lead a life of their own choosing. Since states of nature could be desirable in each of these ways, it follows that they underwrite three reasons for action that people have even in the state of nature. People have a reason to lead a life of their own choosing, they have a reason not to interfere with anyone’s leading a life of their own choosing, and they have a reason to help make sure that everyone has the wherewithal to lead a life of their own choosing. More succinctly, people have reasons to help but not interfere, and apart from that, they have a reason to do whatever they want to do. Even more succinctly, people have a reason to do whatever they want so long as they show each other respect.

I said that any answer to the question of what people have reasons to do is going to be controversial, and this answer is no exception. But at least the answer just given is familiar, as it is in the same ballpark as the answer reached by moral philosophers committed to rationalism during the Enlightenment (see for example Kant 1948 [1786]). According to this Enlightenment view, there are both moral and non-moral reasons for action. If we think of moral reasons as unconditional reasons for action—that is, as reasons that we have independently of our desires—then the answer just given tells us that we have moral reasons to help but not interfere, and that
we also have non-moral reasons to do whatever we want. Though some philosophers reject this answer, one attraction of it is that it is the answer to which those who start out thinking that we should give an answer more like the one Hume gave will find themselves committed at the end of the day \(\text{[Hume 1968 [1740]]}\).

As I understand their view, Humeans begin from the plausible thought that what agents have reason to do is fixed by what they would want themselves to be doing if they weren’t vulnerable to reasoned criticism, but they then add to this the implausible thought that all that reasoned criticism can amount to is criticism of an agent’s beliefs as false or unjustified, on the one hand, and criticism of their desires for failing to connect up with such beliefs in an instrumentally rational way, on the other (see \text{[Williams 1980]}). It is because of this combination of plausible and implausible thoughts that they give the very different answer that agents have only the non-moral reason for action posited by the Enlightenment view: that is, a reason to do whatever they want to do on condition that their wants aren’t based on ignorance or error or instrumental irrationality (here I echo the responses to Williams given by \text{[Korsgaard 1986]} and \text{[Hooker 1987]}).

Let’s agree that what agents have reason to do is fixed by what their desires and beliefs would be like if they weren’t vulnerable to reasoned criticism. This is equivalent to what their desires and beliefs would be like if they fully exercised all of the capacities that underwrite their being agents, where the distinctive feature of agents is that they have the capacity to know what the world in which they live is like and to satisfy their desires in that world (\text{[Smith 2013]}). Agents who fully exercise all of the agential capacities thus know what the world is like, and they satisfy their desires in it, and — and this is the important point that Humeans overlook — they do so robustly. It is no mere fluke that agents who have and fully exercise all of the agential capacities have this knowledge and satisfy their desires. Given that agents are temporally extended, the robust exercise of agential capacities must itself be temporally extended; at the limit, agents who exercise their agential capacities robustly must exercise them at each moment they exist. It follows from this that such agents must have the wherewithal within themselves, or the worlds in which they exist contain the wherewithal, to overcome two kinds of vulnerability to which they would otherwise be subject.

The first kind of vulnerability is the vulnerability of an agent’s later self to their earlier self. Imagine an agent at a certain time who exercises the capacity to know what the world is like and
satisfy their desires at that time. For their exercise of this capacity to be robust, it cannot be the case that that exercise is dependent on the fact that, at an earlier time, they just so happened not to want to deceive themselves at the later time, or that at the earlier time, they just so happened not to want to undermine their satisfaction of their desires at the later time. Moreover, if their possession of those capacities is itself the product of their efforts to develop and maintain their capacities at some earlier time, then it couldn’t be dependent on the fact that, at that earlier time, they just so happened to want to help make sure that they develop and maintain such capacities to exercise at the later time. They must instead have some feature, in virtue of being an agent who exercises all of the capacities that underwrite their being an agent at each moment they exist robustly, that explains why at the earlier time they helped develop and maintain their having their agential capacities, and why at the earlier time they didn’t interfere with their later exercise.

The obvious feature to imagine them having at the earlier time is a pair of desires, constitutive of being an agent whose exercise of their agential capacities is robust, a desire to help ensure that that they have agential capacities to exercise, on the one hand, and a desire not to interfere with their exercise once they have them on condition that their exercise won’t itself constitute interference, on the other. Another obvious feature to imagine them having is the self-control required to make sure that they act in accordance with such desires in nearby worlds in which they are subject to apathy or weakness of will. Agents who lack these two desires, or who lack the capacity for self-control, are thus subject to reasoned criticism.

The second kind of vulnerability is the interpersonal analogue of the intrapersonal vulnerability just described. Imagine an agent at a certain time who exercises the capacity to know what the world is like and satisfy their desires in it at that time. For their exercise of this capacity to be robust, it cannot be the case that that exercise is dependent on the fact that, at that time, there are other agents who just so happen not to want to deceive them, or that at that time, there are other agents who just so happen not to want to undermine their satisfaction of their desires. Moreover, if their possession of those capacities at that time is the product of the efforts of other agents to help them develop and maintain their capacities, it couldn’t be dependent on the fact that those other agents just so happened to want to help them develop and maintain their capacities. Those other agents must instead have some further feature that explains why they help but don’t interfere.
What this suggests is that we would be wrong if we supposed that agents could robustly exercise their agential capacities in worlds in which other agents have just any old desires, and that we were wrong when we earlier restricted the desires that agents have to have to overcome their intrapersonal vulnerability to desires that concern themselves. For an agent’s exercise of their agential capacities to be robust, they must be in the company of other agents who robustly exercise their agential capacities too, and they must all desire to help develop and maintain the agential capacities of every agent, and desire not to interfere with the exercise of the agential capacities of any agent on condition that their exercise wouldn’t itself constitute interference. Moreover, all such agents must have the capacity for self-control required to make sure that, in nearby worlds in which they are subject to apathy or weakness of will, they act in accordance with such desires anyway. Agents who lack these two desires or the capacity for self-control are subject to reasoned criticism.

The upshot is as advertised. If you begin with the Humean idea that what agents have reason to do is fixed by what they would want themselves to be doing if they acted in ways that avoid reasoned criticism, then you will inevitably be led to reject the Humean view that agents have a reason to do just one thing, namely, whatever they can do that will in fact result in the satisfaction of their desires, without regard to what their desires are. The problem with this Humean view, to repeat, is that it cannot explain why an agent is subject to reasoned criticism if their exercise of their agential capacities is fragile rather than robust. We must instead suppose that what agents have reason to do, no matter what their circumstances, is whatever they would want themselves to do in those circumstances in the nearest worlds in which they have and exercise all agential capacities robustly.

If we call agents in such worlds ideal agents, then what makes agents ideal is the fact that they have, in addition to all of the agential capacities and whatever desires they happen to have, a desire to help develop and maintain the agential capacities of all agents and a desire not to interfere with the exercise of the agential capacities of any agent on condition that that exercise wouldn’t itself constitute interference. Moreover, their agential capacities must include the capacity for self-control. An ideal agent’s possession of this capacity explains why, in the nearest non-ideal worlds in which they are subject to apathy or weakness of will, they get themselves to
do what their ideal counterparts desire them to do anyway, which helps explain why their exercise of their capacity to do what they have reason to do is robust.

This is why I said earlier that even those who begin thinking that the Humean view of reasons for action is correct will end up rejecting that view in favor of the Enlightenment view. According to the revised Humean version of the Enlightenment view, everyone has a reason to do whatever they would want themselves to do if they exercised their agential capacities robustly, which is just an elaborate way of saying that they have a reason to lead a life of their own choosing (a non-moral reason); that they have a reason to help make sure that everyone has agential capacities to exercise, which is just an elaborate way of saying that they have a reason to help make sure that everyone has the wherewithal to lead a life of their own choosing (a moral reason); and that they have a reason not to interfere with anyone’s exercise of their agential capacities on condition that that exercise doesn’t itself constitute interference, which is just an elaborate way of saying that they all have a reason not to interfere with anyone’s leading a life of their own choosing so long as their leading that life doesn’t interfere with anyone else’s leading a life of their choosing (another moral reason). Indeed, we might well think that our discussion of the Humean view explains why the Enlightenment view is correct. For while the Humean is right that it is a conceptual truth that what agents have reason to do is fixed by what their desires and beliefs would be like if they weren’t vulnerable to reasoned criticism, the Enlightenment view provides the best account of the substance of our reasons for action, as there is an argument that takes us from the Humean’s conceptual truth to the Enlightenment account of the substance. Since agents could act in each of the three ways just specified in the state of nature, the question to which we must ultimately turn is whether they favor the creation and maintenance of social rules. However, before turning to that question, three more points need to be made about the three reasons.

The first is that there is indeterminacy in the content of the three reasons. Consider the reason not to interfere with anyone’s exercise of their agential capacities, by way of example—similar points could be made about the other two reasons. If you cause someone to be in extreme agony, then they may be so overwhelmed that they are unable to act at all. But now imagine a continuous sliding scale of pains with extreme agony at one end and only the faintest feeling of pain at the other, a feeling so faint that it doesn’t interfere with anyone’s exercise of their agential capacities, and imagine a range of actions that cause the pains that lie at each point along
this sliding scale. At what point do these actions amount to interference with an agent’s exercise of their capacity to know what the world is like and realize their desires in it? The answer is that there is no such point. Instead there is a range of actions that are determinately instances of interference, and a range of actions that are determinately not instances of interference, and in between there is a range of cases about which it is indeterminate whether they are cases of interference.

The second point that needs to be made concerns the weights of the three reasons. To stick with the example of the reason not to interfere—again, similar points could be made about the other two reasons—this reason will be more or less weighty depending on the extent to which the interference in question undermines an agent’s exercise of their agential capacities. Given that the three reasons can conflict with each other, this means that there is no easy way to spell out the relative weights of the three reasons in cases where they conflict. To return to the less elaborate way of spelling out the content of the reasons, the option best supported by all three reasons taken together will be that option in which the mix of living a life of one’s own choosing, not interfering with anyone’s leading a life of their own choosing, and making sure that everyone has the wherewithal to lead a life of their own choosing, is that which is most similar to the mix in the ideal case. In some cases, the moral reason to help will outweigh the moral reason not to interfere, in other cases vice versa. In some cases, the moral reasons to help but not interfere will outweigh the non-moral reason to lead a life of one’s own choosing, and in other cases vice versa. Moreover, since there is bound to be vagueness in this similarity relation, the weights that these reasons have will themselves almost certainly be vague too.

The third point that needs to be made about the three reasons for action agents have is that, given the point just made about the weights of the three reasons, we can provide the following plausible definitions of the deontic statuses of actions. Actions that are morally wrong are those there are decisive moral reasons not to perform; actions that are morally permissible are those that there are no decisive moral reasons not to perform; and actions that are morally obligatory are those that there are decisive moral reasons to perform. Since we have also seen that non-moral reasons may sometimes outweigh moral reasons, we can also define the following novel deontic statuses: actions that are non-morally wrong are those there are decisive non-moral reasons not to perform; actions that are non-morally permissible are those that there are no
decisive non-moral reasons not to perform; and actions that are non-morally obligatory are those that there are decisive non-moral reasons to perform.

With these three points in mind, let’s return to the task of assessing the normative significance of social rules. The question to which we want an answer eventually is whether it is ever permissible or obligatory, either morally or non-morally, to create social rules in the state of nature and then maintain those rules. Given the three points just made, this will evidently be an extremely difficult question to answer. In the first instance, let’s therefore ask an easier question. Are there either moral or non-moral reasons to create social rules in those states of nature in which agents both know what reasons for action they have and are robustly motivated to act on those reasons, social rules that there would then be reasons to maintain? Since this is simply a state of nature in which all agents are ideal, we can rephrase the question. Would ideal agents in the state of nature want themselves to create and maintain social rules?

3. Ideal Agents and Social Rules

Since ideal agents robustly have and exercise maximal capacities for desire realization and knowledge acquisition, it might be thought that in a state of nature in which all agents are ideal social rules would have no role to play. But this would be a mistake for two reasons.

The first is that ideal agents will still need to coordinate their behavior with each other, and that social rules are a means by which they can coordinate. For example, imagine a state of nature that includes a beautiful view from a mountain top that only one person can occupy at a time, and imagine that all of the ideal agents in that state of nature desire to take in that view. In order to satisfy their desires, given their desires to help and not interfere, they will have to find a way to stagger their climbing of that mountain, and if taking in the view is more central to the life plans of some, then they will also have to take into account their level of interest. This is a coordination problem, and a social rule to the effect that (say) people wishing to visit the mountain book the day and time of their visit in advance, with the amount of time that they can spend there being a function of how central taking in the view is to their life plan, would solve the problem. We can therefore readily imagine ideal agents in a state of nature like the one just described exiting that state of nature in favor of a world in which some such social rule exists. Their non-moral reasons to satisfy their desires as efficiently as possible provide them with reasons to establish and maintain the rule, and their moral reasons to help and not interfere
ensure that the rule they establish and maintain will not place unfair burdens on anyone who is subject to it.

The second reason we can readily imagine social rules being created by ideal agents in a state of nature is connected to two of the features of the three reasons identified above. We saw that the content and weight of these reasons is vague. But in order to act on these reasons in the full range of circumstances in which ideal agents find themselves, they would need much more precise accounts of what it is to help and not interfere, and what the relative weights of these reasons are vis-à-vis each other and vis-à-vis doing they want. Though in their interactions with their intimates individual ideal agents could find themselves jointly committing to some more determinate conception of these reasons without too much thought or planning, there is no reason to expect that different ideal agents and their intimates would commit to the same more determinate conception. When it comes to interactions with strangers, some means of converging on a more determinate conception of the content of the reasons and their weights would therefore be required, and here we again see a role for social rules. The behavior of others, as affirmed by those in the sub-group, would provide the means by which the more determinate conceptions could be communicated to those who aren’t already in the know.

So far, we have focused on the reasons ideal agents have to create and maintain social rules, thereby exiting the state of nature, that take as given that they are and will remain ideal. One of the questions we had about social rules is whether they are only significant because of the possibility of our failing to recognize what we have reason to do, or of our failing to respond to our recognition of that fact by being motivated accordingly, or whether they would still be significant even if we didn’t fail in one of these ways. We now have our answer to that question. Social rules would still be significant even if we didn’t fail in one of these ways, as they would have a role to play even in possible worlds in which all agents are ideal. A further observation about the role of social rules in the lives of ideal agents is also worth making. Given that ideal agents can have contingent desires about anything, they may well have contingent desires about how they are all to coordinate, and about which more precise conception of what it is to help but not interfere that they all converge upon. Since their reasons to help but not interfere may not provide them with a decisive reason not to attempt to change existing social rules so that they better align with their contingent desires—provided, of course, that they maintain the existence of social rules that facilitate what’s fully determinate in the pre-social ideas of helping and not
interfering—we would expect there to be a good deal of debate among ideal agents about what the social rules are to be going forward. It would therefore be unsurprising if social rules, even in the ideal case, were in a state of flux.

We should therefore resist the temptation to think that a possible world in which everyone knows what their moral reasons are and is motivated accordingly would be a world in which there is a total harmony of interests. Given the reasons that there are to create social rules, depending on what contingent desires ideal agents happen to have, there could well be conflict, albeit good-natured conflict that would be resolved peacefully and in which some people’s current satisfaction with the social rules would be conditional on the possibility of their changing them in the future (compare Strawson 1961). Moral concepts, even in the ideal case, thus turn out to be ‘essentially contested,’ to use W.B. Gallie’s helpful term (Gallie 1956), because their extension is fixed part by social rules that are what they are in part because of the outcomes of civil arguments over what they are to be.

4. Ever-So-Slightly-Non-Ideal Agents and Social Rules

Let’s now relax the assumption that we are in a state of nature in which all agents are ideal, but only ever-so-slightly. Given that ideal agents are only contingently ideal, the best way to do this is by imagining ideal agents contemplating the contingency that they will become less ideal. Would ideal agents want certain sorts of social rules to be created and maintained in that contingency?

Imagine a state of nature in which agents have less than maximal capacities for desire realization and knowledge acquisition, but in which they still have the wherewithal to figure out what their ideal counterparts want them to do in their world, and the self-control to get themselves to do it. For example, adapting an example of Scanlon’s, imagine two such agents hunting on opposite sides of a deep river, one with his boomerang and the other with his spear (Scanlon 1998). They each launch their weapon at their prey, miss, and their weapon sails across the river to the other side, landing within easy reach of the other. Because weapons are difficult to make and use in the circumstances we’re imagining—boomerang-throwers can’t easily make or accurately throw spears, and spear-throwers can’t easily make or accurately throw boomerangs—each of them is keen to get their own weapon back, but neither of them can swim. They therefore begin making a complicated series of gestures to convey the idea that, if the other
throws their weapon back across the river to them, they will do the same in return. The boomerang-thrower eventually throws the spear back. The question is whether the spear-thrower has a reason to throw the boomerang back.

When Scanlon discusses this case, the question he asks is slightly different from the one that interests us. He asks whether it would be **wrong** for the spear-thrower not to throw the boomerang back, and he tells us that it would be wrong because the following principle, which he calls ‘Principle F,’ is true:

```
If (1) X voluntarily and intentionally leads Y to expect that X will do such-and-such (unless Y consents to X’s not doing so); (2) X knows that Y wants to be assured of this; (3) X acts with the aim of providing this assurance, and has good reason to believe that he or she has done so; (4) Y knows that X has the beliefs and intentions just described; (5) X intends for Y to know this, and knows that Y does know it; and (6) Y knows that X has this knowledge and intent; then, in the absence of special justification, X must do such-and-such unless Y consents to such-and-such’s not being done.
```

*(Scanlon 1998, p. 304, emphasis added)*

Principle F tells us that even though it might not be wrong for X to fail to create the reasonable expectations he creates in Y, it would be wrong for X to fail to meet these reasonable expectations, having created them.

A striking feature of the fact that everyone has moral reasons not to interfere with each other’s exercise of their desire-realization and knowledge-acquisition capacities—that is, the ideal counterparts of the non-ideal agents imagined would want them not to interfere—is that the following modified version of Principle F is also true (the changes are emphasized):

```
If (1) X voluntarily and intentionally leads Y to expect that X will do such-and-such (unless Y consents to X’s not doing so); (2) X knows that Y wants to be
```
assured of this; (3) X acts with the aim of providing this assurance, and has good reason to believe that he or she has done so; (4) Y knows that X has the beliefs and intentions just described; (5) X intends for Y to know this, and knows that Y does know it; and (6) Y knows that X has this knowledge and intent; then, in the absence of special justification, X has a moral reason to do such-and-such unless Y consents to such-and-such’s not being done.

According to the modified principle, even though our moral reasons not to interfere may or may not tell in favor of our creating certain expectations, they do tell in favor of meeting them once we have created them. This is because someone who knowingly or negligently creates expectations in someone, and then fails to meet those expectations, interferes with that person’s exercise of their capacity for both knowledge acquisition and desire realization. They knowingly or negligently lead them to falsely believe that they will meet the expectation, and thereby cause them to act on desires that go unsatisfied in circumstances in which, if they hadn’t been deceived, they would have acted so as to satisfy different desires instead.

Note that the modified version of Principle F is not itself a social rule. The spear-thrower in the state of nature is in a position to reason himself to the conclusion that this principle is true simply by imagining what his ideal counterpart would want him to do, thereby reaching the conclusion that he has a moral reason to return the boomerang. But the complicated series of gestures that the two hunters have to go through in order to make the antecedent of Principle F true in the state of nature provides them and other members of their social group with reasons of efficiency to create a social rule that allows for the creation of those expectations more swiftly, perhaps with a simple gesture, or the making of a noise such as, ‘I hereby promise to do such-and-such.’ To be sure, the expectations thereby created will be those codified in the social rule, and these mightn’t be exactly the same as the expectations that agents would have created if they had been willing to go to the trouble of making all of the complicated gestures in a one-off interaction. In deciding whether to institute the social rule, agents will therefore have to trade off the efficiencies created by reliance on the social rule against the benefits of going it alone. But the fact that the efficiencies will oftentimes win in such a trade-off suffices to prove the point, namely, that it is plausible to suppose that the ideal counterparts of the two hunters would want
social rules about promising to be created and maintained, and that this may well be what they most want.

The example of the two hunters is merely illustrative. In that case, the expectations created track fairly closely the expectations created in cases of promising. But the social rules that would be instituted for reasons of efficiency, given the truth of the modified version of Principle F, wouldn’t be restricted to social rules about the creation of those expectations. If we considered a different state of nature case in which two agents knowingly create in each other certain expectations about who gets to do what with certain objects in their respective possession, then reasons of efficiency will plausibly tell in favor of creating social rules that codify property rights. Once again, since the expectations thereby created will be those codified in the social rule, and since these mightn’t be exactly the same as the expectations that agents would have created if they had been willing to go to the trouble of making all of the complicated gestures, in deciding whether to create the social rule governing the control of objects, agents will have to trade off the efficiencies created by reliance on the social rule against the benefits of going it alone. But the fact that the efficiencies will oftentimes win in such a trade-off once again suffices to prove the point. The ideal counterparts of less-than-ideal agents in a state of nature who possess and control property would want social rules governing property to be created and maintained as well, and that this too may be what they most want.

5. Even-More-Non-Ideal Agents and Social Rules

We have so far focused on whether ideal agents and ever-so-slightly-non-ideal agents have reasons to exit the state of nature by creating certain social rules. The crucial feature of these cases has been that the agents know what they have reason to do and are motivated to do it. Let’s now consider state of nature cases in which some agents are even less ideal. They either have false beliefs about what they have reason to do and are motivated to act on these false beliefs, or they know what they have reason to do but aren’t motivated to act on their knowledge. The ideal counterparts of these agents will want them to exit this state of nature by creating social rules that have a different character from those we’ve considered so far.

Imagine a bully in the state of nature just described. When someone is in his way, he pushes them out of the way, so acting contrary to his reason not to interfere. What do those who get pushed out of the way have reason to do? Since their reason not to interfere is itself
conditional—it is a reason not to interfere with others on condition that they are not themselves interfering—this reason does not tell in favor of their not interfering with the bully. But since they also have a moral reason to help make sure that everyone has capacities for knowledge acquisition and desire realization to exercise, the bully included, it follows that they don’t have carte blanche to interfere with him in any way they like either. They have a non-moral reason to do something to stop him from interfering with them, grounded in their non-moral reasons to satisfy the desires that the bully is an obstacle to their satisfying, and that might include interfering with him, but they also have moral reasons to leave his capacities for knowledge acquisition and desire realization intact.

Here are some examples of actions that fit this description: getting the bully to stop by conveying to him that he is doing something he has a moral reason not to do, and then relying on his exercise of self-control to get himself motivated not to do it; not engaging in any voluntary associations with the bully until he stops pushing them around; and physically constraining the bully in a way that doesn’t escalate the situation. Consider the first of these. If the bully is open to the evidence that he has a reason not to push someone around, and the self-control required to get himself motivated to act on that knowledge having acquired it, then the interference will stop. But these are big ‘ifs,’ as in the non-ideal state of nature we are imagining it isn’t obvious that these conditions will be met. The bully might be belligerent and so immune to reason; he might possess the capacity for self-control, but fail to exercise it; he might not care whether he has any voluntary association with the person he pushes around, given that he can just push them around; and the power differential between them might make it impossible for the bully to be physically constrained by the one he is pushing around. So far, then, it isn’t clear how those being interfered with can get those who interfere with them to stop.

Fortunately, third parties in the state of nature also have reasons not to interfere with anyone on condition that they aren’t themselves interfering, so this reason doesn’t tell in favor of their not interfering with the bully either. Moreover, though they too have reasons to help ensure that everyone has capacities for knowledge acquisition and desire realization to exercise, and hence don’t have carte blanche to do whatever they like to the bully, these reasons have special relevance for the person the bully is pushing around. This is because though the existence of power differentials means that those being interfered with may not be able to deal with those doing the interfering themselves, those power differentials disappear when third parties are
involved, as third parties can act as part of a group. This fact about the superior power of third parties acting as part of a group, suggests that the preparedness of third parties to step in and prevent those like the bully from pushing people around is itself plausibly partially constitutive of everyone’s having knowledge acquisition and desire realization capacities to exercise. This is because unchecked interference itself constitutes a diminution of people’s capacities.

The reasons third parties have to help therefore provide them with reasons to step in, perhaps as part of a group, to stop the bully from interfering. As before, their doing so is constrained by the need to make sure that the bully still has knowledge-acquisition and desire-realization capacities to exercise. Examples of actions third parties could perform as part of a group that fit this description include: getting the bully to stop by conveying to him that he is doing something he has reasons not to do, and relying on his self-control to get himself motivated not to do it; making it clear to the bully that none of the other members of the social group will engage in any voluntary associations with him until he stops; and physically constraining him. When performed by third parties acting as a group, these actions are much more likely to be successful. Though the bully’s belligerence might make him immune to reason from someone he is bullying, it is less clear that he would be immune to reason from an entire group, at least if he has the foresight to realize that the members of the group will be disinclined to engage in voluntary associations with him if they think he can’t be relied on to do what he has reason to do. And, if all else fails, the group will certainly have the power to physically constrain him. So, once we think about third parties acting as a group, it becomes clear how those being interfered with can get those who interfere with them to stop. They can get them to stop by getting third parties involved as a group, something that third parties acting as a group have moral reasons to do.

Here is another role for social rules. Given that the preparedness of third parties acting as a group to step in and prevent interference is itself plausibly partially constitutive of everyone’s having knowledge-acquisition and desire-realization capacities to exercise, everyone will have both moral and non-moral reasons to create social rules that make that preparedness clear to everyone. Everyone will have such reasons because the mere existence of the social rule may itself suffice to dissuade those who are inclined to interfere not to interfere. To the extent that this happens, everyone gets the benefit of the preparedness of third parties to step in and prevent interference without having to pay the price of being one of the third parties who is called upon
to do so. Of course, these social rules will therefore have to codify the nature of the response of groups of third parties to interference, and these responses may not have the nuance that could be achieved by forgoing social rules and having groups of third parties step in on a case-by-case basis. Those in the state of nature will have to trade off the efficiencies created by the existence of the social rule against the benefits of remaining in the state of nature and stepping in on a case-by-case basis. But the mere fact that the efficiencies will oftentimes win in such a trade-off once again suffices to prove the point. The ideal counterparts of those in such circumstances would plausibly want such social rules to be created and maintained in such circumstances, and in some cases it may be what they most want.

6. Actual World Agents and Social Rules

Let’s finally ask about the significance of actual social rules. What does what we have learned about the reasons agents in the state of nature have for exiting the state of nature and creating social rules tell us about the reasons for action that we have in the actual world?

Let’s start with what it doesn’t tell us. Though the fact that our ideal counterparts in the state of nature desire that we in the actual world help but don’t interfere, and apart from that do whatever we like, shows that in the actual world we have reasons to help but not interfere, and apart from that to do whatever we like, the fact that our ideal counterparts would immediately exit the state of nature and create certain social rules does not show that we have reasons to abide by those social rules. Those social rules are tailored to the circumstances of ideal agents, and our circumstances are nothing like those. Moreover, we already live in a world in which there are social rules, and these are social rules we have reasons to abide by and maintain to the extent that they enable us to coordinate; to the extent that they give more determinate content to the abstract ideas of helping and not interfering, and apart from that doing whatever we like; to the extent that they constitute institutions of promising and property; and to the extent that they codify the preparedness of third parties to step in and prevent interference. The fact that they are not the social rules that our ideal counterparts would want created and maintained in the nearest world in which we are in a state of nature is neither here nor there.

Since the social rules that exist in the actual world didn’t come into existence because of the actions of agents who were all acting on their reasons to help and not interfere, and apart from that to do what they like, the extent to which they have these features that make them defensible
will be partial at best. We therefore have reasons to engage in a delicate balancing act. We have reasons to maintain existing social rules to the extent that they give determinate content to the ideas of helping and not interfering and apart from that our doing what we like to which others freely sign on, while at the same time modifying them so as to make them more fully defensible: that is, so as to make them have determinate content to which more others will willingly sign on, or to which others will more willingly sign on. Modifying them in these ways is an especially delicate task given that we do not have the ability to change social rules all by ourselves, and given that, even if we did have that ability, our reasons to help and not interfere would tell against our doing so. What the reasons to help and not interfere tell in favor of is our changing the social rules by getting others who are subject to them to agree to those changes. Note that this need not require us to get others to agree that we all have reasons to help and not interfere, and apart from that to do whatever we like. It only requires us to get others to agree, for whatever reasons they might happen to think that there are, to changes that the reasons to help and not interfere favor being made to the existing social rules. This would involve trying to find common ground with others, and then arguing for changes to the social rules based on that common ground (compare Rawls 1993).

The upshot is that what actual agents with knowledge of what they have moral reasons to do have knowledge of is a compromise whose boundaries are not static. What counts as a compromise shifts with changes in contingent facts about the numbers of other people with whom they have to find common ground, and the amount of common ground they can find. If this makes actual agents’ acquisition of knowledge of what they have moral reason to do sound like what they have knowledge of is the outcome of a successful political negotiation in advance of that outcome, then that is no accident. For what we have learned is that, given the crucial role played by social rules in giving determinate content to what we have moral reasons to do, acquiring knowledge of what we have reason to do is a lot more like engaging in a successful political negotiation than we might have thought it would be. The contrast between this account of what it is to acquire moral knowledge and that offered by normative ethicists who claim that the canonical way to acquire moral knowledge is individualistic—a matter of getting our intuitions about specific cases and moral principles into a reflective equilibrium with each other—couldn’t be more stark.
Notwithstanding this contrast, it is striking that this account of what we have knowledge of when we have knowledge of what we have moral reasons to do, suggests that some of the most urgent moral problems we face are those that currently command so much grassroots attention. One of the most important tasks we face in getting political negotiations off the ground is securing the recognition of everyone affected. Similarly, one of the most important tasks we face in modifying existing social rules is ensuring that those rules give due recognition to everyone whose ability to live a life of their own choosing is affected by the social rules under which they live. It is therefore heartening that the exclusionary nature of so many existing social rules—exclusion based on sex, sexual orientation, race, ethnicity, religion, wealth, physical and mental abilities, and species—is the subject of so much current debate (see Calhoun [2000] for a suggestive defense of the norms of civility that constrain us when we engage in such debate).

References


