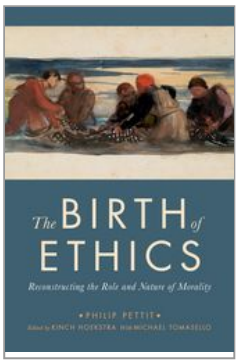

The Birth of Ethics: Reconstructing the Role and Nature of Morality
Philip Pettit and Kinch Hoekstra

Print publication date: 2018
Print ISBN-13: 9780190904913
Published to Oxford Scholarship Online: October 2018
DOI: 10.1093/oso/9780190904913.001.0001

Ground Zero

Philip Pettit

DOI:10.1093/oso/9780190904913.003.0004

Abstract and Keywords

Ground zero is the condition in the society addressed—“Erewhon,” in a familiar acronym of “nowhere”—before the advent of moral or ethical concepts and practices. In this society, we, the inhabitants, use natural language to make reports to one another on ourselves and our environment and, being want a reputation for telling the truth, we generate a norm of careful and truthful reporting. But reports may be excused in two epistemic ways: by invoking a misleading or a changed environment. And it is striking that, while we may report on our attitudes—our individual, internal environments—we do not avow or pledge them. An avowal would foreclose the misleading-mind excuse, as we human beings might be expected to be able to do in view of our alleged capacity for self-knowledge. And a pledge would foreclose the changed-mind excuse as well, as might seem to be possible in view of our alleged capacity for self-control.

Keywords: Norm, truth, reputation, self-knowledge, self-control, excuse, avowal, pledge

This chapter is meant to introduce the society of Erewhon at a point where members have access to natural language but use it only for exchanging basic information about their environment. The characterization of this starting point naturally falls into positive and negative parts. The positive looks at what has to be present in the beginning of the story presented here, exploring how members of Erewhon use natural language in the exchange of environmental information and looking then at the effects this use is likely to generate. The negative offers a sketch of some speech acts that are saliently absent at this point of the

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

narrative, in particular the committal speech acts of avowing and pledging that will assume prominence in the third and fourth chapters.

Language and communication

The natural language that we in Erewhon use at this initial stage of development —at ground zero—is designed to achieve communication of a relatively sophisticated kind. In the normal case, such communication involves the intentional and manifest transmission of information, or what I as speaker take to be information. To rework one standard analysis of the main conditions, I use my words with the primary intention of conveying that information to an audience and with the secondary intention of achieving that result, at least in part, by making my primary intention salient (Sperber and Wilson 1986; Grice 1989).

Strictly, even these two intentions can be present without an act being communicative in the fullest sense associated with language. Consider a case in which I want you to refill my glass of wine, put the **(p.58)** glass where you cannot help but notice that it is empty, and fully expect that you will realize that I want it filled and respond to my desire. Or consider a case illustrated beautifully in a poem of Coventry Patmore's (Quiller-Couch 1922, 1023).[1] The heroine mimics sleep in the hope that her lover will kiss her without her seeming, immodestly, to invite it. She is charged with immodesty by an imaginary interlocutor but has a ready, if ingenuous response:

> "I saw you take his kiss!"
>
> "*Tis true*."
>
> "O, modesty!"
>
> "*Twas strictly kept:*
>
> *He thought me asleep; at least, I knew*
>
> *He thought I thought he thought I slept*."

Do I communicate in the original case that I want a glass of wine? Does the heroine communicate that she wants a kiss? In both cases, arguably not. In the first example, I do not want it to be overt between us that I am seeking (yet) more wine; in the second, the heroine does not want it to be overt that she is seeking a kiss. Each of the agents wants to maintain a façade, in the one case of indifference to alcohol, in the other of sexual modesty. For the full communication associated with speech, I must not hide any of my intentions from you in this way: I must keep them overt, as it may be said (Neale 1992, 550; Moore 2017). It may even be that in full communication the intentions have to be manifest to both of us, being a matter of common awareness or common

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

ground: each of us is in a position to be aware of the intentions, in a position to be aware that each is aware of them, and so on in the usual hierarchy.[2]

(p.59) These complexities need not be of further concern here, since the argument can go through under any of a range of accounts. But it is important to recognize that they are in place and that they put Erewhon on a par in that respect with familiar human communities. It is complexities of this kind that distinguish communication in human language, or so at least it seems, from the transmission of information by the signaling systems used among other species (Scott-Phillips 2015).

In this chapter, the story will first sketch the alignment and cooperation we are bound to achieve—the norms we are bound to generate—as speakers who communicate after this fashion in reporting on our common world and, by extension, our attitudes toward it. Then, in a second section, it will identify other modes of communicating our attitudes, involving avowals and pledges, that are absent at this stage. And then in a final section, it will look at reasons why avowing and pledging are not as hazardous as they may at first seem: why the availability of practical excuses, and exemptions, makes them feasible.

The narrative that begins here will continue in the third and fourth chapters, explaining why we will have the means, the motive, and the confidence to go beyond mere reports of attitude and commit ourselves in avowals and pledges, individual and shared. And then in the fifth and sixth chapters, it will show how the practices of avowal and pledging can make concepts like those of the morally desirable and responsible available to us and give them a role in the regulation of our conduct.

The narrative that begins in this chapter and continues to the end of chapter 6 will be presented from the viewpoint that I and you and others purportedly occupy as residents of Erewhon; in this narrative, then, ‘we’ will generally refer to us in this imagined role. Chapter 7, like chapter 1, will shift the viewpoint back from the narrative that we in Erewhon occupy to the viewpoint we all share in thinking about ethics. With the **(p.60)** reconstructive analysis in place, it will return to a consideration of what that analysis teaches about the moral world that we ourselves occupy.

2.1 On what is necessarily present

Beyond free-riding

The reconstruction of money begins with a purely barter society in which the members perform the most elementary form of in-kind trade, exchanging goods and services with one another. The reconstruction of ethics begins with a linguistic community in which members use language in an equally elementary role, exchanging information about their environment with one another. They

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

exchange information about aspects of the shared world on which the speakers are presumptively more informed than the audience.

Thus, I may report to you, as you may report to me, that the berries on the hill are ripening. Or you may report to me, as I may report to you, that the weather up north is getting better. There is no problem about explaining why we should each be interested in this mutual exchange of information. If I know about the berries and you do not, then you will benefit from learning that they are ripening; this will direct you to a potential source of nutrition and pleasure. If you know about the weather up north and I do not, then I will benefit from hearing that it is improving; that will open up the opportunity for a journey that I might otherwise have thought impossible.

When we take one another's words at face value, believing what we are told, then we rely on one another for the information, or would-be information, that they convey. And that reliance will be personally beneficial and appealing for each of us insofar as the other is a reliable informant.

But while it is going to be attractive for each of us to be able to rely on the words of others, it need not always be attractive to provide information on which others can rely with benefit. The benefit I give you in telling you about the ripening berries comes at a cost to me, since **(p.61)** I could have had all the berries to myself had I said nothing or had I misinformed you: had I told you, for example, that they would not be ripe for another week or so.

Am I likely to be tempted in such a case to free-ride: to rely on the information you give me about the weather but to deny you information about the berries? I may be exposed to the temptation, but I am unlikely to succumb. We in Erewhon, like people everywhere, live our lives in continuing interaction with many of those with whom we have exchanges. And that means that I am unlikely to be disposed to misinform you about something like the berries. For if I misinform you about such a matter, you are unlikely to rely on my words on future occasions and you are unlikely to prove reliable in later interactions, or at least in any interactions in which telling the truth would impose a cost on you. Thus, if I misinform you about the berries, I run the risk of suffering various costs. I am likely to lose the ability to rely on you when I need some information you can provide and, equally, I am likely to lose the ability to get you to rely on me when it is in my interest that you do so.

This is going to be so, at any rate, under certain assumptions about the situation that are likely to be generally satisfied. They include the specific assumptions, first, that in most informational exchanges, as in the one described, the benefits of deception are not important enough to outweigh the costs of being seen to misinform you; and, second, that there is no way of keeping you in the dark, say by equivocation or a pretense of ignorance, that allows me to escape those costs.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

Equally, they include the more general assumption that the society is small enough, and that I am recognizable enough, to make it impossible for me in most cases to dodge detection if I do misinform you.

The focus in the discussion that follows will be on those cases in which there is no easy way for a deceiver to avoid detection, the expected costs of detection are high in relation to the benefits of deception, and there is no easy way of keeping you in the dark without deceiving you. By assumption, such cases are statistically normal in Erewhon.

As it will be unattractively costly for me to prove unreliable in dealing with others in such exchanges, so it will be unattractively costly for you, or indeed for anyone else in Erewhon; it will reduce your ability to rely **(p.62)** on others and to get them to rely on you. Thus, it will be a profitable strategy for each of us to make a habit of proving to be informative and reliable reporters. We may seek to get away with lying occasionally, even in normal cases, but it will generally make most sense to prove reliable.

This lesson is reinforced by a further consideration. Suppose I lie to you about the berries. Not only will I be exposed to the possibility of being unable on future occasions to rely on you or to get you to rely on me; I will also be exposed to the possibility that you will spread the word about my unreliability. It will be in your interest to tell others about my failure, after all, since that information will help you to prove yourself reliable to them. And for parallel reasons it will be in the interest of each of them to pass on the word to their interlocutors. Thus, it may even become a matter of common awareness in the relevant circles that I misled you.

In dealing both with you and with others, then, most acts of deception are liable to trigger a serious reputational setback, leading to my being cast as someone with whom it is hazardous to do business. The prospect of such reputational costs is going to be salient for any one of us in Erewhon, and the desire for reputation is likely to dispose each of us to be speakers whom others can depend upon to tell the truth.

Assuming that I lie occasionally, and only occasionally, may I expect to benefit reputationally from any instances in which I tell the truth? Not in general; and not, in particular, when word gets around about my failures. My performance will tend to generate a general reputation that attaches to me as a person, for two distinct reasons.

One is that it would be unnecessarily difficult for others to record and remember how I behaved, now in this circumstance, now in that, now with this interlocutor, now with another; and unnecessarily difficult for others to form situation-specific expectations about how I am likely to deal with them in particular, or with them in a new context. It will be much easier for them to give me a general

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

label as a person, attributing a general disposition to tell the truth reliably or a general disposition of indifference toward the truth.

(p.63) The second reason for this reputational focus on the person goes back to "a candidate for the most robust and repeatable finding in social psychology," as E.E. Jones (1990, 138) describes it. This, known as the fundamental attribution bias, is "the tendency to see behavior as caused by a stable personal disposition of the actor when it can be just as easily explained as a natural response to more than adequate situational pressure." The presence of that bias in human beings means that if I ever tell lies in Erewhon, then I am in danger of being dubbed a liar; and that I must reliably tell the truth, if I am to be treated as a habitual truth-teller. In the first case, my behavior will be explained by an imputed indifference to the truth, in the second by an imputed disposition to tell the truth reliably.[3]

The fact that reputation focuses on people across different situations, and not on situation-specific performances, is going to raise the reputational stakes enormously. Since this focus is bound to be a matter of common awareness, it means that I or you or anyone else in Erewhon will face serious costs of a kind that it is hard to undo, assuming detection and publicity, if we fail to tell the truth to others. It means that we will live within an economy of esteem, where our truth-telling status in the eyes of the community is on the line in almost every informational exchange (Brennan and Pettit 2004).[4]

(p.64) The norm of truth-telling

If we are to be reliable informants and win out reputationally in Erewhon, then we must prove ourselves reliable in two distinct respects. First, we must be careful in processing information about any issue on which others query us: we must consider all the relevant data and take them properly into account. This is going to be attractive for us independently, in any case, since our own welfare will often depend on getting things right. Second, we must transmit that information truthfully or sincerely to our audience, communicating that things are as we find them to be. Putting these requirements together, the lesson is that we will generally be motivated to be competent or careful observers on the one side and sincere or truthful speakers on the other.

To the extent that in normal cases we each generally manage to prove reliable on these two fronts, truth-telling will become a general pattern or regularity in Erewhon. It will emerge as a result, now in this case, now in that, of our being careful about processing any information sought and truthful about transmitting it. But it may become a general pattern of that kind without our noticing that it is a feature of our society. It may emerge, without our being aware of it, behind our backs. It may be an unforeseen, aggregate consequence of how we are each disposed to behave in our individual exchanges.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

Under the story told, there is also another general pattern or regularity that will characterize our society. Other things being equal, we will each be aware in any interaction that by proving reliable and telling the truth, we can win a favorable reputation with our audience and with anyone who learns of our performance. We will each expect, case by case, that by telling the truth we can win the good opinion—or at least, avoid the bad opinion—of those who are directly or indirectly involved.

Thus, there will be a regularity of attitude established in the society—a pattern in our expectations—that goes along with the behavioral regularity of telling the truth. As in the behavioral case, this regularity may also escape our notice in Erewhon. No matter how resiliently established, it may be an aggregate consequence of our individual responses that we do not register as a social fact.

(p.65) Apart from these behavioral and attitudinal regularities, there is also a third sort of regularity—if you like, an explanatory regularity—that is likely to materialize among us. Not only will I and you and others be each generally disposed to tell the truth in making reports to others, and not only will we be each disposed to expect a reputational payoff from truth-telling. It is also likely to be the case that our disposition to tell the truth will be reinforced, now on this occasion, now on that, by that expectation: it will be boosted by the recognition that it may be reputationally hazardous, now in this exchange, now in that, to mislead our interlocutor. And this may be the case on more or less every occasion, of course, without our registering that it holds as a matter of general regularity; as in the other cases, the regularity may escape our notice.

There are different ways, it should be noted, in which my expectation of reputational rewards may support my telling the truth. It may be a sufficient motive for telling the truth, even the only motive that weighs with me. It may be a motive that supplements a distinct motive to tell the truth, where only the complex of incentives is sufficient to support truth-telling. Or it may play yet a third role if I already have a sufficient motive for telling the truth, and do so as a matter of habit or virtue. In this case, the interest in reputation or esteem can serve to reinforce the independent disposition to tell the truth, being there to keep me on the truth-telling track should that motive or habit fail; in a word, it may serve as a backup to counter the possibility of failure.

The observations made about the general pattern of truth-telling displayed in statistically normal informational exchanges can be summed up in three clauses.

- Almost everyone in the community conforms to the regularity of telling the truth, at least when dealing with those who have not proved deceptive in the past.
- Almost everyone expects such conformity to attract a favorable reputation among others and/or nonconformity to attract an unfavorable reputation.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

(p.66) • Almost everyone is supported in their conformity to the truth-telling regularity by the expectation of such reputational benefits and costs.

The sort of regularity that truth-telling exemplifies is not a mere convention, such as driving on the right- or left-hand side of the road (Lewis 1969). Such a convention, unlike the regularity of truth-telling, is grounded in the attraction for each of us of doing the same thing as others. Thus, the convention does not give rise to free-riding problems of the kind that truth-telling occasions. You will not be tempted in approaching another car to break convention, drive on the same lane, and risk a head-on collision. But absent reputational costs you may be tempted to tell me a lie while relying on me to tell the truth; if you think detection is unlikely, you may well try to deceive me (Ullmann-Margalit 1977).

The three conditions it satisfies mean that the truth-telling pattern is distinct, not just from mere conventions, but also from a number of other social regularities. That it is generally maintained in people's practice, as the first condition stipulates, distinguishes it from a standard honored more in the breach than in the observance, such as a supererogatory ideal that we routinely applaud but rarely realize. That it reflects the mutually expected attitudes of inhabitants toward conformity, as the second condition holds, means that it is distinct from a regularity to which others are manifestly indifferent, such as the regularity whereby most people sleep at night, not during the day. And that it is supported by that expectation, as in the third condition, means that it is distinct from a regularity such as taking steps to guard against penury in old age; it is unlikely that people are motivated in any degree by expecting that others will think well of them for displaying such prudence.

Unlike a convention, then, the truth-telling regularity in Erewhon is not supported just by the wish to do the same as others. And unlike the remaining rivals, it is a pattern that is generally realized in the practice of the society, implicated in the prevailing attitudes of members, and supported, case after case, by the expectation of satisfying those attitudes.

(p.67) A pre-social norm

This sort of regularity is often described as a norm, albeit with variation in how exactly the conditions are understood. It attracts general conformity among the population and does so insofar as it is supported by the expectation of their having positive attitudes to conformity, and negative attitudes to non-conformity.

As there is almost certain to be a norm of truth-telling in Erewhon, so there is equal reason to expect that there will be similar norms against violence, fraud and other failures of reciprocal beneficence. The manifest interest of members in the reliability of others, and the attraction for others of satisfying that interest, explains why there is likely to be a norm of truth-telling. And the manifest interest of members in the beneficial reciprocity of others, and the attraction for others of satisfying that interest, explains why norms of these

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

distinct sorts are likely to materialize as well. Those other norms will not figure explicitly in the narrative that now follows, but they will figure at a later point as examples of what we in Erewhon, having developed moral concepts, are likely to take as desirable patterns.

A norm in the sense elaborated so far does not constitute a criterion that guides us in our choices: a pattern that we intentionally track. It does not constitute a rule in the sense of being a criterion "consulted by those whose behavior is being assessed" (Brandom 1994, 64). For all the narrative supposes, we may not even be aware of the norm that we bring into existence, let alone take it as a pattern to try to realize in our behavior. For this reason, the sort of norm involved may be described as pre-social only; more in a moment on what may give it the status of a rule to consult, transforming it into a social norm.

It makes sense to say that we inhabitants of Erewhon regulate or police one another into conforming to a pre-social norm like telling the truth. But it is important to realize that this need only be an unconscious and unintentional mode of regulation. It may exist solely by virtue of the fact, on the one side, that we each look to how reliable any interlocutor is, letting our experience dictate how we treat the person in future and how we testify about them to third parties; and, on the **(p.68)** other, that in most cases we each seek to prove reliable as interlocutors ourselves, seeing this as the price to be paid for favorable treatment and testimony by others.

As it makes sense to speak of mutual regulation for truth-telling in the Erewhon characterized so far, so it also makes sense to say that we, the members of the society, cooperate with one another in generally telling the truth. But just as mutual regulation of the kind envisaged may be unconscious and unintentional, so the same is true of the cooperation we can be said to practice. The reason is that under the psychology postulated, each of us may think of what we do in telling the truth merely as a means to gaining or preserving a personal, reputational benefit, not as a means of realizing an aggregate end that is to everyone's advantage. While we align our behavior with one another in telling the truth, we each act for our own ends; we do not necessarily act out of a shared or cooperative intention to play our part in securing a social goal.

These points are meant to emphasize the fact that the truth-telling society of Erewhon, as it has been established so far in the narrative, is very simple. The cooperative patterns into which we members regulate one another do not require an awareness of the regularities or an intention to conform to them. They emerge by a reputational motor whose effects—benign effects, in the case of a norm like truth-telling—are not necessarily visible to us; they appear, in Adam Smith's famous phrase, as by an invisible hand.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

A social norm

But is a pattern like general truth-telling likely to remain unnoticed in Erewhon? And is it likely to remain incapable, therefore, of guiding us in the manner of a rule? Or is it going to achieve that recognition and attain that role only at a point, which comes much later in the narrative, where we in Erewhon gain access to ethical concepts?

There is reason to think that the pattern will become visible to all before we develop such concepts, and indeed that it will become visible to all as a matter of common awareness, so that everyone is aware of the **(p.69)** visibility, everyone is aware that everyone is aware of it, and so on. The reason to think that this is plausible is, first, that the evidence of the regularity in truth-telling is going to be available to all, thereby making it visible; second, that it is also going to be evident to all that this evidence is universally available, thereby making it visibly visible; and third, that as that is true in the relation between the first and second level, so by an inductive step it is going to hold in the relation between the second and third, between the third and fourth, and so on (Lewis 1969).

Assuming a degree of perspicacity on the part of members, then—and a tendency for perspicacious observations to spread—it is plausible that at some point members will come to share in a common awareness of the truth-telling regularity they sustain. And the same will hold, presumably, for any similar norm against violence or fraud or whatever.[5]

This inductive argument for common awareness of the behavioral regularity involved in the norm of truth-telling, or any similar norm, applies also to the other regularities involved: to the attitudinal regularity according to which everyone expects a reputational sanction for how they behave and to the explanatory regularity according to which this expectation supports conformity in everyone's case. This means that in all likelihood any norm like that of truth-telling will satisfy a fourth condition over and beyond the three already listed:

- Almost everyone believes, as a matter of common awareness, that the first three conditions are satisfied.

If truth-telling satisfies this common-awareness condition, thereby assuming the status of a manifest norm, then sooner or later, that is likely to trigger a certain chain of consequences. And those consequences, which we proceed to review, will make it into a social norm, proper—that is, a norm that can serve for us as a guiding rule.

(p.70) Manifestly, any member of the society can see the norm, once it is recognized in common awareness, as a pattern such that their full acceptance within the society depends on conforming to it. Manifestly, any member can then communicate that observation to any outsider seeking entrance, to any child at the point of social induction, or indeed to any erstwhile or would-be offender,

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

themselves included. And so, manifestly, the norm of truth-telling—and, by extension, any similar norms like those against violence or fraud—can present as a rule of behavior that each must take as a guide, if they wish to belong properly to the community; indeed, it can present as a norm that we each have an interest in encouraging others to accept and abide by.

When norms of truth-telling, non-violence, non-fraudulence or whatever get established in this way as rules, they count as properly social norms. While they may continue, at bottom, to depend for their maintenance on the engine or motor of reputational pressure, conformity to those norms can now appeal on a new count: that the norms direct us to paths we must follow if we are to remain in good standing within the society. While continuing to rely on reputational pressures for motivational support, the norms can also appeal on this social basis.[6]

With this development, the regulation and cooperation that we in Erewhon exercise over one another in eliciting truth-telling or any **(p.71)** similar pattern will no longer be as blind as it would have been prior to the realization of the fourth, common-awareness condition. But it will still not be a form of prescriptive, let alone moral regulation, for it will not depend on the employment of any concept of desirability. We may treat the social norms of our society as appealing without having any sense of them as desirable norms. Indeed, for all that has been said, we may not even have any idea of what it might be for something to count as desirable. The norms may identify a social payoff for conformity, without giving us a prescriptive attitude towards them.

2.2 On what is saliently absent

Beyond self-reports

To the extent that we are anxious to prove ourselves reliable to others, we must have an interest not just in reliably conveying how things are in our shared environment, but also in reliably conveying that we are reliable speakers. We must want to convey, for example, that we form beliefs as the data and only the data require, that we desire or intend to communicate how things are according to those beliefs, and more generally that we have only desires and intentions of a kind that would make us congenial parties in interaction.

How can we convey this information? How can we communicate about ourselves rather than our surroundings: about our internal psychology rather than our external environment? We can certainly make reports on ourselves in the way in which we make reports on the world around us. But are there any other ways in which we might communicate to others the shape of our personal dispositions, thereby reassuring them about our credentials as parties with whom they can do business?

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

On the face of it, there are two salient possibilities. In terms of art, which will be explained shortly, these are: on the one hand that we might *avow* some attitudes rather than reporting them; and on the other that we might *pledge* certain attitudes rather than avowing or **(p.72)** reporting them. These possibilities become visible, as we shall see, in light of ways in which they vary two striking features of mere reports.[7]

In setting up the contrast between avowing or pledging an attitude on the one side and merely reporting it on the other, there is no suggestion that we are likely in Erewhon to go through a temporal stage at which the established practice is to report on our attitudes to one another without any avowal or pledge. Consistently with the points to be made here, and with the narrative developed in coming chapters, avowing and pledging attitudes may come more naturally to us than reporting them. The point of the contrast is to mark the distinguishing features of avowals and pledges and, as will appear in the next chapter, to show that those features can make it appealing to opt for an avowal and a pledge where a report would have been at least an abstractly possible and acceptable alternative.

Two features of reports

In order to be a reliable reporter on the environment, speaking the truth about how things are, I must process information reliably and I must transmit information reliably. But it will be obvious to those of us in Erewhon, as it will be obvious to any creatures of the human ilk, that I may falter in either of these exercises, while still remaining someone on whom you can generally rely as an interlocutor.

Putting aside practical problems in transmitting information—more on these later—there are two salient epistemic problems that may affect the processing of information. These epistemic problems would explain my failure to tell you the truth but do so in a manner that argues against your keeping me at a distance, refusing to rely on me. If they are accepted as explanations of my failure on a specific occasion, they will save my general reputation as a reliable truth-teller.

The first sort of problem that might serve in this role involves misleading evidence. It will be obvious to all of us in Erewhon, as it will be **(p.73)** obvious in any human society, that the evidence on which I or anyone else relies in forming and reporting a belief may be misleading. I may have been quite sure that the berries on the hill were ripening. But still I may have been misled. For it may have been that I only had the chance to see them late in the day. And the berries may have looked red and luscious in the fading sunlight, when earlier observation would have shown that they were still relatively green.

In this first type of problem, my words fail to match the world as it was at the time I observed it: my report about the ripening berries did not fit the actual

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

facts. In a second type of failure, my words may have matched the world at the time of the observation but they convey the message that it continued that way afterward when, as a matter of fact, it didn't. In this failure, the world fails to stay matched to those words: the facts observed and reported cease to obtain. Thus, it may be in the case of the berries that a third party went and picked the ripe berries before you made your way to the hill and judged my report inaccurate. The world I observed did not mislead me but it changed or altered between the time of my observation and the time of your acting on my report of that observation.

Under standard psychological assumptions, which we in Erewhon may be expected to endorse, my failure to tell the truth about the berries in either of these cases is due to a factor over which I had no control. It was not because of having been careless in processing information—or, presumptively, in transmitting it—that the report I gave you about the berries turned out to be false. Rather it was due to the world having let me down. In the first case, the world presented itself in a misleading manner. In the second, it changed between the time of the observation that led to my report and the time when you acted on that report.[8]

These observations show that even while we in Erewhon sustain a social norm of truth-telling, reliably communicating about how things **(p.74)** are in our environment, we may see good reason not to impose the usual reputational costs on a speaker who fails to tell the truth. We will suspend those costs, and continue to treat the speaker as reliable, when there is a plausible misleading-reality or changed-reality explanation for the failure. It would not make strategic sense for us to take any other line, since doing so would close down lots of opportunities for profitable, future exchanges; it would eject people unnecessarily from the network of mutual reliance.

From epistemic excuses to avowals and pledges

This being so, it should be clear that when I prove to have made a misreport, I will be happy if I can explain my failure in one of these ways. If you accept the explanation that I offer, or that someone else offers on my behalf, then you will have every reason from your own point of view to overlook the failure and to continue to relate to me as someone who can generally be depended upon to speak the truth. The explanation will let me off the reputational hook and keep our relationship as mutually reliable and reliant speakers in place.

In letting me off the hook in this way, the two explanations count as excuses for my not having told the truth. They are epistemic excuses insofar as they cite problems in the processing of information rather than problems in its transmission: problems that thwart my best efforts to be careful in recording how things are rather than my truthfulness in conveying how I take them to be.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

As will appear later, other sorts of excuses cite practical problems in the transmission of information to the same reputation-saving effect.

While such epistemic and practical problems may let me off the reputational hook, counting functionally as excuses, they are not excuses that presuppose ethical or moral ideas. They presuppose the social norm of truth-telling, and the reputational discipline that keeps that norm in existence among us. And we may naturally grade them for their explanatory potential in a way that presupposes a sense of the epistemically plausible. But they do not require that we in Erewhon recognize truth-telling as desirable or that we hold one another morally responsible for **(p.75)** telling the truth. If they did, they could not figure at this point in the narrative.[9]

Either of the two epistemic problems envisaged in the example with the berries would provide a full explanation of my misreport, not just a partial one, and let me entirely off the reputational hook. It is possible to imagine other problems that would go only some way toward explaining that departure from truth-telling, however, and go only some way toward letting me off the hook. These might be cast as partial rather than full excuses. In this narrative, the category of excuses—practical as well as epistemic excuses—will be limited to full excuses alone. This will make the story much simpler than it might have been. And yet it should not damage the purpose that the narrative is meant to serve; it will be clear as things proceed that room can easily be made for introducing partial as well as full excuses.

Not only does the narrative focus only on full excuses; it also focuses only on excuses of a purely synchronic character. My being intoxicated may well provide a synchronic, misleading-world excuse for having misread and misreported the ripeness of the berries on a given occasion. But it may not let me off the hook over time. It may be clear that I should be treated as an unreliable processor of information by virtue of the fact that, prior to the observation, I was careless about maintaining my information-processing capacity: I let myself get drunk. This shows that a synchronic excuse will let me off the hook only if it is also satisfactory on the diachronic front: only if it is not undermined by a past performance for which there was no synchronic excuse at the time.

Simplicity argues for keeping this complication out of the narrative, however, and for treating synchronic excuses as the only relevant category; and it argues for this line with practical as well as epistemic excuses. Like the focus on full excuses alone, this simplification need not compromise the purpose of the narrative. It ought to be clear in this **(p.76)** case too that room can always be made for acknowledging that excuses impose diachronic as well as synchronic requirements.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

This account of the epistemic ways in which a misreport can be excused directs us to the two modes of communicating things that count as avowals and pledges. These voluntarily and manifestly foreclose appeal to epistemic excuses in communicating that something is the case. To communicate that something is the case—say, that you hold a certain attitude—while foreclosing appeal to the misleading-reality excuse counts as avowing that that is the case; to communicate that it is the case while foreclosing appeal to both the misleading-reality and the changed-reality excuse counts as pledging that it is the case.

It is a particularly strong requirement on avowal and pledging that they have to be voluntary. I do something voluntarily only if I do it intentionally: that is, roughly, as a means of satisfying my desires according to my beliefs. I may do something intentionally, however, without doing it voluntarily; think of what I do in handing over my money to the mugger. In order to do something voluntarily, it must be that as I saw things—rightly or wrongly—there were alternatives to the intentional action that promised to be broadly acceptable.[10] In requiring that avowal and pledging be voluntary, the idea is that the speaker chooses without pressure to foreclose the misleading-reality and the changed-reality excuses that reporting would leave in place.

(p.77) Self-access and the salient absence of avowal

The fact that any misreport can be excused after the event by appeal to the misleading character of the relevant domain is bound to be a matter of common awareness: something of which we are each conscious, each conscious that each is conscious, and so on. But that means that in making a regular report in Erewhon, I will do so in a way that manifestly keeps open the possibility of appealing to a misleading reality in excusing a misreport: a misleading world in the case of a misreport about independent facts, a misleading mind in the case of a misreport about my own attitudes. Any such report will embody caution, providing me with possible safeguards against proving to have told an untruth and being consequently ejected from the ranks of those reputed to be reliable.

A reporter’s caution is well placed when I communicate about our shared world, or about the world of another’s mind, for that reality is an elusive domain about which it is manifestly easy for me to be misled. But things are intuitively different when it comes to communicating about my own mind. There is a long tradition of thinking that I have much more intimate access to my mind than I have to the external world. And if that access is secure enough with a given attitude, so it would seem, then I ought to be able to communicate about that attitude without the same caution: without keeping open the possibility, in the event of a miscommunication, of claiming to have been misled about the attitude conveyed. In our term of art, I ought to be able to avow the attitude rather than just report it.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

The intimacy of my contact with myself, so the idea goes, would enable me to communicate that I have the attitude—say, a belief or desire or intention—while making it manifest that should I prove later to have misspoken, I will not seek to get off the reputational hook by appeal to having been misled about my mind. In making an avowal, in this sense, I will not present myself as a mere reporter on my mind—that is, as someone who could later excuse a misreport by saying that appearances were deceptive. Rather I will present myself as being in a position to speak about my attitudes without any possibility—certainly without any real-world possibility—of being mistaken.

(p.78) Avowing an attitude is a speech act that would voluntarily and manifestly set aside a reporter's caution. It would foreclose the possibility of appealing to the misleading character of my mind in excusing a failure to display the attitude that I communicate. It would reflect a special degree of confidence in the self-ascription of the attitude: an assumption that in speaking about it I enjoy the authority of a privileged spokesperson.[11]

Self-control and the salient absence of pledging

There is no avowal of attitude in Erewhon, as the society has been characterized so far. And neither is there any pledging. The idea of pledging an attitude—say, pledging an intention—mirrors the idea of making a pledge, absent unforeseeable obstacles, to behave in a certain manner; it amounts to making a pledge to behave in a way that reflects the presence of that attitude. Where avowing an attitude would voluntarily and manifestly foreclose the possibility of appealing to a misleading-mind in excusing a miscommunication, pledging an attitude would go one better. It would voluntarily and manifestly foreclose the possibility of appealing to any epistemic excuse, whether of the misleading-mind or changed-mind variety.

The notion of making a pledge to behave in a certain manner is close to that of promising to act in that way. But the pledging of attitude envisaged here does not involve promising in a moral or ethical sense of the term. In that sense, to promise would be to engage or create a moral obligation, other things being equal, not to break the promise. In the sense in which pledging is introduced here, it need have no such connotation. As avowing would foreclose the possibility of getting off the reputational hook by appeal to having been misled about my mind, pledging would also foreclose the possibility of getting off the reputational hook by appeal to having had a change of mind. The crucial point to notice in each case is that the hook in **(p.79)** question is reputational in character, and does not involve any idea of moral or ethical censure.

It is understandable that I might claim to know my mind sufficiently well to be able to avow an attitude—this, because of the assumption of self-access—without being sure enough about likely changes of mind to be able to pledge it. Thus, I might claim to know enough about my intention—say, an intention to join you on

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

a hunt—to be able to support an avowal, and yet not know enough about how my plans are likely to change to be able to make a pledge. I might feel that I could assure you of the intention, being quite certain of its existence, without feeling able to make a pledge to go to the event, thereby guaranteeing to continue to have the intention.

Why might the idea of pledging an attitude make any sense, then? And why might its absence from Erewhon be salient? The reason is that by long tradition I not only have better access to my own mind than to the shared world; I also have control over my own mind of a kind that I do not have over our shared world. Assuming that I have special control over some of my attitudes—say, my intentions—as well as special access to those attitudes, it makes sense to think that I might be able to set aside both of the epistemic excuses that reporting the attitudes would keep in place. In the presence of such access and such control, the idea is that I could foreclose the possibility of excusing a miscommunication about an attitude like an intention in either of the two ways that reporting would leave open: by claiming that my mind had been misleading or by claiming that I had changed my mind.

If I avowed an intention to join you on a hunt, as in an earlier example, then I could try to excuse my failing to turn up, and my disappointing you, by the claim that I changed my mind. All that would be precluded is an attempt to explain and excuse the failure by claiming to have been misled about my own mind. But if I pledged the intention, assuring you in that extra manner that I would turn up, then I could not invoke that excuse either. I would have foreclosed both the misleading-mind and the changed-mind excuse and would have claimed to speak authoritatively for myself, not just on the basis **(p.80)** of a special form self-access but on the basis of a special self-control as well.

2.3 Excuses and exemptions

Are avowing and pledging feasible?

Avowing and pledging are two sorts of speech act that, by the account presented so far, are saliently absent from Erewhon. We members of Erewhon use language only in the most basic way imaginable as a means whereby to make reports to one another about the world and perhaps even about our own minds. We do not avail ourselves of the communicative options—specifically, options in communicating our attitudes—that would appear to be open as a result of our special access to our own minds and our special control over them.

But let it be granted in principle that we in Erewhon might make avowals or pledges of attitude, despite our not actually doing so—that our self-access and self-control should make this possible. There is still a question as to whether avowals and pledges would represent feasible options for us to take. Would it not be prohibitively risky for me, or for any one of us, to set aside excuses in the manner that they require? Would it not be utterly rash to try to meet their

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

requirements, reducing my ability to explain a miscommunication in a face-saving manner—that is, in a way that would preserve my reputation as a reliable interlocutor?

Various considerations that are relevant to this issue will come up in the next chapter. But for now, it is important to emphasize that the risk apparently associated with avowing and pledging is not as great as it may seem. For even if I foreclose appeal to one or both of the epistemic problems that might save my reputation in the event of misspeaking, I may still be able to appeal with plausibility to certain practical problems or to certain forms of exemption, as it is often called, in order to excuse a miscommunication. Unlike the epistemic problems, these difficulties all involve factors such that I will never be in a position to foreclose the possibility of appealing to them as explanations of a miscommunication.

(p.81) Practical excuses, and exemptions

The practical problems that might be invoked in excusing a miscommunication affect the transmission of information rather than its processing. Unlike epistemic problems, they jeopardize, not my carefulness in observing how things are, but my truthfulness in conveying how I take them to be.

There are a variety of factors that might explain why I give a misleading account of how things actually are, despite having processed the facts correctly. One problem might be that pressure of time or company leads me to speak without thinking properly and to blurt out a falsehood. And another might be that I am coerced or induced not to tell the truth for fear of the consequences—perhaps my career is at stake or someone has put a gun to my head. Were you persuaded that I spoke falsely as a result of any such problem, then you might well conclude that you should not eject me from the network of those you can rely on.

Practical problems arise in a distinctive way with the transmission of information about how things will be as a result of an intention I avow or pledge. In this case my truthfulness may be affected, not via an obstacle to my saying how things are, but via an obstacle to my causing things to be as I say they will be. Thus, suppose that I make a pledge to meet up with you at sundown but fail to turn up at the agreed meeting place. The fact that I broke a leg, or was forcibly kept at home, will provide a practical excuse for what may seem like my untruthfulness or insincerity in communicating that intention.

The fact that I broke a leg or was forcibly detained will get me off the reputational hook in this way because of two natural and manifest assumptions. First, that I as speaker could not plausibly have taken that possibility into account when I avowed or pledged my intention. And second, that you as audience could not plausibly have taken me to have made the avowal or pledge in a manner that allowed for the obstacle. You could only have taken me to

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

communicate the intention, and predict or pledge the behavior, on the assumption that no hindrance of that open-textured type—no unforeseen obstacle—would get in my way.

(p.82) Epistemic excuses focus on something that goes wrong in my processing of information and in my effort to be careful about what I report. Practical excuses direct attention to problems or limitations that affect my transmission of information, undermining my effort to be truthful in what I say. While both may assume the form of partial excuses rather than excuses of a complete sort, and while both may be required to satisfy a diachronic as well as a synchronic requirement if I am to be truly off the hook, the focus in the evolving narrative will be on complete, synchronic excuses alone. This focus will make the presentation easier, as noted already, without leaving more complex excuses shrouded in mystery; the amendments required for admitting them should be fairly clear in the different cases discussed.

Epistemic and practical excuses both invoke problems that impede my capacity to use words reliably, whether in making reports, avowals or pledges; both explain why, despite having misspoken, I should still be treated as a reliable interlocutor. But apart from such excuses, it is important to recognize a third category of face-saving explanation that might be offered for a failure to prove reliable, whether in making a report, an avowal, or a pledge. This is the exempting explanation that would suggest that at the earlier time of utterance—or, if relevant, the later time on which the utterance bears—I was not fully adult or able-minded: I was not a functional interlocutor or agent (Watson 1987; Wallace 1996; Gardner 2007).

An exempting explanation might cite the hypnotic influence under which I spoke in my original utterance, a bout of paranoia that affected what I said, or a moment of compulsion that overcame me in trying to live up to my words. It would suggest that I was in some sense out of my mind and not fit to be held to the normal expectations that operate under the discipline of mutual reliance. It would argue for keeping me in the ranks of those reputed to be reliable, at least if the impairment cited ceases to remain a presence or a prospect—if I clearly get over the effect of the hypnosis or paranoia or compulsion.

The idea here is not, as in the case of an excuse, that the exercise of my general capacity to prove reliable was impeded by a problem in the processing or the transmission of information. Rather it is that that **(p.83)** capacity itself was impaired, whether for a time or over an extended period, and that I should be exempted from being put on the hook for the associated failure. Such an exemption, like an excuse, may be partial or total, and it may or may not allow me to be put on a diachronic hook. But as in the case of excuses, the focus throughout this study will be exclusively on complete, synchronic exemptions.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

Looking ahead

By the account offered, avowals and pledges are saliently and perhaps surprisingly absent in the Erewhon described up to this point in our narrative. Their absence is surprising because of the intuitive idea that we have a special form of access to our own minds that would make avowal possible, and a special kind of control over what we do that would make pledging possible.

Since this project is essentially naturalistic, nothing in the narrative that follows can assume that we have a form of access or control that would be scientifically inexplicable. Thus, the narrative must explain the emergence of avowals and pledges, and of the concepts of desirability and responsibility, without appealing to a non-naturalistic access to the self of the sort that Descartes postulated or to a non-naturalistic control over the self of the kind posited by libertarian accounts of free will.

There is no need to appeal to such ideas, because it turns out that there is a compelling, naturalistically intelligible reason why we inhabitants of the society should be driven to enrich the communication of our attitudes to include both avowals and pledges. The explanation of why we should be pushed in this direction is the topic of chapters 3 and 4, which address the commitments constituted by avowing and pledging attitudes, on the one side, and by co-avowing and co-pledging them on the other.

The notion of commitment introduced in those chapters, needless to say, does not have a prescriptive sense. It does not presuppose that we in Erewhon have access to any prior idea of desirability or responsibility but only that we are capable of betting on ourselves to live up to the words we utter in avowing and pledging attitudes.

(p.84) This notion of commitment is commonly invoked in game theory and economics and applies clearly to avowals and pledges. These are acts of communication in which speakers voluntarily expose themselves to certain costs that they must bear in the event of failing to speak the truth. In effect, they constitute wagers in which speakers bet on themselves to prove reliable, accepting that they will lose their stake if they fail. They will be unable to access a misleading-mind excuse if they fail to live up to an avowal, and unable to access either a misleading-mind or a changed-mind excuse if they fail to live up to a pledge.

Chapter 3 argues that, starting from Erewhon as it has been characterized so far, we would be individually likely to make such commitments in our own name. And then chapter 4 argues that we would be equally likely to make commitments in the name of others as well as ourselves; in other words, we would be likely to make commitments with others as well as commitments to others.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

Notes:

(1.) Cited in Blackburn (1984, 110). I reorganize the lines of the poem, in order to distinguish clearly between heroine and interlocutor.

(2.) The hierarchy need not be problematic, imposing the impossible requirement that everyone should have formed a positive view on the question that arises at each level: whether everyone is aware, is aware that everyone is aware, is aware that everyone is aware that everyone is aware, and so on. All that need be the case is that each is disposed for any question that may arise at any level—assuming the capacity to understand it—to provide a positive answer. Here and throughout I presuppose the more or less standard account of common awareness or belief, usually associated with Lewis (1969), setting aside any issues about its proper articulation. I assume that whatever revisions may be required, they will not impact on the argument of the book. See Lederman (2018) for an overview of background issues.

(3.) The finding that people are deeply prone to the fundamental attribution bias means that, even when they are conscious of their own sensitivity to reputational considerations (Miller and Prentice 1996, 804), people will be loath to trace the behavior of others to such a situational pressure.

(4.) The pattern envisaged here is somewhat different, then, from the pattern of tit for tat characterized and illustrated by Robert Axelrod (1984). Under tit for tat, we will each be disposed to tell the truth to anyone we are dealing with for the first time and to treat anyone else as they treated us in the immediately preceding interaction. If I and you each tit for tat, the upshot will be that we each speak the truth to one another on first encounter and, absent a failure on either side, continue to do so in all future encounters. The esteem-based mechanism is consistent with such tit-for-tat effects but not reducible to them. Generalizing tit for tat beyond two-party interactions brings it closer to the economy of esteem, but has problems of its own: see Pettit (1986).

(5.) As noted previously, I presuppose the more or less standard view of common belief throughout this book and, as the text indicates, I also presuppose a more or less standard view of when common belief is likely to emerge. Whatever amendments are required in a more exact account of common belief and its emergence, they are unlikely to impact on the overall argument. For background, see Lederman (2018).

(6.) For an earlier version of this conception of a social norm, see Pettit (1990) and Brennan and Pettit (2004); the current version appears in Pettit (2008b, 2015c). It is more fully elaborated in Pettit 2019b. This notion of a social norm picks up points made in a variety of approaches. See for example Hart (1961); Winch (1963); Coleman (1990; Sober and Wilson (1998; Elster (1999); Shapiro (2011). For an insightful development of the idea of reputationally supported

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020

norms, see Appiah (2010). And for an overarching theory that is reconcilable with that adopted here, although it uses terminology somewhat differently, see Brennan, Eriksson, Goodin and Southwood (2013). Notice that a convention in Lewis's sense may also count as a norm, satisfying the four conditions given, although it need not do so; the conditions that make it a convention do not suffice in themselves to make it a social norm in our sense.

(7.) For an independent style of argument in favor of the role of such initiatives in social life, see Zawidzki (2013, Ch 7).

(8.) The focus here is entirely on empirically elusive, logically contingent claims. Suitable adjustments will be needed for reports on self-evident or necessary truths. See the short discussion in the last section of chapter 3 of framework beliefs.

(9.) For an excellent account of the nature and role of excuses in moral thinking generally, see Kelly (2013).

(10.) This account is modeled on that in (Olsaretti 2004), although there are differences between us; one is that on that assumed here, it is possible to do something involuntarily but willingly: that is, roughly, with relish. What makes an alternative broadly acceptable? Perhaps, anticipating ideas not yet introduced in Erewhon, that in a suitable context I could be held responsible for not taking it; the alternative does not hurt or offend or challenge me—it does not involve such a cost or such a difficulty—that I cannot be blamed for avoiding it. Why assume that the alternative need only be an apparent option? Because, to turn to a sort of case made famous by Harry Frankfurt (1988)—for more discussion, see chapter 6, section 1—I may voluntarily do X when, unbeknown to me, a third party would have stopped me from taking the alternative option, Y; in such cases doing Y is merely an apparent option.

(11.) For an extended, broadly congenial, account of avowals in more or less this sense, see Bar-on (2004).

Access brought to you by:

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (oxford.universitypressscholarship.com). (c) Copyright Oxford University Press, 2020. All Rights Reserved. An individual user may print out a PDF of a single chapter of a monograph in OSO for personal use. Subscriber: UC - Berkeley Library; date: 22 August 2020