



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

Committing to Others

Philip Pettit

DOI:10.1093/oso/9780190904913.003.0005

Abstract and Keywords

We in Erewhon will have the means, the motive and the confidence to avow many of our beliefs. A means of avowing the belief that *p* is to assert that *p*, since this forecloses the misleading-mind excuse. A motive for doing so is that the communication thereby becomes more expensive and more credible; I deny myself a way of getting off the hook in the event of a miscommunication: viz., by claiming I must have misread my own mind. And the confidence required is supplied by the fact that I can consciously make up my mind on various issues by consulting the data and seeing where they lead me; I do not have to resort to introspective self-scanning. By parallel arguments we in Erewhon will be led to avow not only beliefs but also desires and intentions, relying on desiderata to play a role analogous to that of data in the case of beliefs. And we will be led by the same token to pledge various intentions—and intentions only—foreclosing the possibility of changing our mind as well as the possibility of having been misled about our minds. Avowals and pledges of these kinds are commitments in the non-moral, game-theory sense: they represent wagers in which we stake our reputations on living up to our words.

Keywords: commitment, avowal, pledge, belief, desire, intention

In this chapter, the narrative suggests that having established a practice of reliable reporting, we in Erewhon would find ourselves drawn also into more committal forms of communication. We would be presented with the means of avowing and pledging certain attitudes, not just reporting them; we would have a motive to avail ourselves of that opportunity; and we would have a basis for the confidence that the exercise requires. The chapter starts with the avowal of

belief, then moves to the avowal of desire and intention, and turns finally to the pledging of attitudes.

It will sometimes be useful in the narrative that follows to speak as if there is an initial stage at which we merely report our attitudes to one another and that we move out of that stage as the possibility and appeal of avowing and pledging attitudes become salient. But, as already noted, this temporal sequencing should not be taken literally. The point of the story is just to mark the features that distinguish avowing and pledging from reporting attitudes and to show why they are likely to make avowals and pledges into attractive options.

3.1. The avowal of belief

The means of avowing beliefs

I will avow a belief by the account already given if I voluntarily communicate it in a manner that manifestly forecloses appeal to a misleading-mind excuse for any miscommunication. I voluntarily communicate it in the sense that I convey it in the presence, or at least the apparent presence, of acceptable alternatives. And I do this manifestly insofar as **(p.86)** it is a matter of common awareness between me and my interlocutors that I do so.¹

The possibility of avowing a belief is immediately accessible for any one of us in Erewhon. For it turns out that when I report on any state of affairs, say, a fact about the external environment, I convey information about my belief in that state of affairs at the same time that I convey information about the state of affairs itself. I convey that information, moreover, as I convey information about the environment, in a characteristically communicative manner. And, most importantly, I convey it in such a way that I manifestly foreclose the misleading-mind excuse.

Why would my reporting that the berries are ripening, that the weather is improving up north or that an acquaintance is honest convey, not just that fact, but my belief in the fact? The answer turns on the tight connection between report and belief. To report that something is the case is to purport to be disposed, evidence remaining unchanged, to act as if it is the case: to purport, in effect, to believe that it is the case. If you reported that something was the case and failed to display that disposition or belief, then we would treat that failure as a sign that you should not be taken at your word. Thus, if you do take me at my word when I say that the acquaintance—let's call him Mogli—is honest, then you must take my words to convey both that he is honest and that I believe that he is honest.

Not only do I convey the information about my belief when I say that Mogli is honest; the fact that I convey it will be a matter of common awareness between us, and indeed among any knowledgeable witnesses. This is because the evidence that I convey it will be available to all of us, the evidence that that

evidence is universally available will be available to all of us, and so on in the usual hierarchy (Lewis 1969). When I make the report that Mogli is honest, then, it ought to be the case that we each believe that I purport to believe that he is honest, that we each believe that we each believe that I speak with this purport, and so on.

If it is a matter of common awareness that in reporting the fact that Mogli is honest, I convey also my belief that he is honest, and if I choose **(p.87)** to make that report, then it is intentional on my part—and this, again, as a matter of common awareness—that I convey the belief as well as the fact. At the least I acquiesce in manifestly conveying that belief. I may not desire or intend to convey that message as such but, manifestly, I foresee that it is a side-effect of conveying the fact that Mogli is honest, which I do intend to convey. I intentionally convey the belief that he is honest even if what I intend as such is only to convey the fact of his honesty.

This means that when I say that Mogli is honest I convey my belief in his honesty in a broadly communicative manner. In communicating a linguistic message of any kind, by the analysis mentioned early in the last chapter, I hold—indeed I hold overtly—by two intentions. I overtly use my words with the primary intention of conveying the associated information to an audience and with the secondary intention of achieving that result, at least in part, by making my primary intention salient to them (Sperber and Wilson 1986; Grice 1989). While my explicit intention in the case of Mogli targets the fact that he is honest, it carries over implicitly to the fact that I believe that he is honest. And so, I may reasonably be said to communicate that I believe that Mogli is honest just by communicating that he is honest.²

The difference between these two forms of communication is often cast in terms of a distinction between the semantic and the pragmatic message of an utterance. That Mogli is honest is the semantic message of the words I use; that I believe he is honest is the pragmatic message. While the utterance reports the fact that constitutes the semantic **(p.88)** message, it expresses the fact that constitutes the pragmatic. The semantic message is available to you in virtue of the significance of the words I use in giving my report; the pragmatic message is available in virtue of the significance of my action in making the report.

Why would these observations point the way toward my being able to avow a belief rather than just reporting it? The answer, to come to the crunch point, is that pragmatically or expressively communicating that I believe that Mogli is honest forecloses resort to a misleading-mind excuse for not displaying that belief.

I could invoke a misleading-mind excuse for miscommunicating my belief in Mogli's honesty only if I had presumably reflected on my beliefs—if you like,

scanned my mind—before communicating the belief to you. But the presumption is that I did not scan my mind in that way when I communicated the belief. Rather, assuming that I took care over what I said, I scanned the world for evidence of Mogli's honesty, and in reporting that he was honest I expressed the belief that he is honest. But if I did not rely on scanning my mind—if I did not need to look at the evidence, introspective or behavioral, for whether I held the belief in his honesty—in order to communicate the existence of the belief, then I cannot be excused for having miscommunicated it on the grounds that the mind I scanned was misleading. A problem in that scanning procedure could explain and excuse my miscommunication only if the scanning procedure actually played a role in prompting the communication. And in the pragmatic case, it does not.

If it turns out that I do not display the belief that Mogli is honest in my behavior, therefore, I will not be able to explain the pragmatic miscommunication of that belief by appeal to a misleading mind. I may prove to have been wrong in ascribing honesty to him and may be able to excuse the semantic miscommunication by pointing out that the evidence about his honesty was misleading: perhaps I mistook someone's testimony to Mary's honesty as testimony to Mogli's. But if I prove to have been wrong in expressing a belief in his honesty—if I prove not to have had that belief—I cannot excuse the pragmatic miscommunication by holding that the evidence of my having that belief was misleading. The reason is that while I may have reflected on the evidence of Mogli's honesty before **(p.89)** communicating semantically that he is honest, I clearly did not reflect on independent, say introspective, evidence about my mind before communicating pragmatically that I believe that he is honest.

By this analysis, and by evidence from the familiar world, it would be laughable to respond to the evidence of such a miscommunication by saying that it seemed clear to me that I held that belief but I was misled by the evidence about my state of mind. This would be laughable because, manifestly, evidence would not have played the sort of role in pragmatic or expressive communication that would allow me later to claim that it misled me.

Even under the narrow characterization of Erewon offered in the first chapter, then, it should be clear that I and others will have been put in a position where it is possible to avow beliefs. Suppose you ask me whether I believe Mogli is honest. And assume that I am in a position where I might report that I have that belief. I might make an utterance to the effect that I apparently hold that belief, keeping open the possibility of excusing myself, should it turn out that I do not actually have it—should it turn out that I do not act as if he were honest. If I reject that perfectly acceptable alternative and choose to respond to your query by expressing the belief—reporting simply that Mogli is honest—then I must count as avowing the belief. I communicate the belief in a way that forecloses the misleading-mind excuse. And since reporting it instead is an acceptable

alternative, I do so voluntarily, not out of any sort of necessity or under any sort of pressure.

The motive for avowing beliefs

Sooner or later we Erewhonians are bound to recognize that communicating a belief by expressing it allows of only one of the excuses that reporting the belief would have permitted and that it exposes us to an increased reputational loss. Might we be tempted in that case to play safe and choose to report our ground-level attitudes of belief about any matter rather than avowing them? Rather than expressing the ground-level belief that Mogli is honest, thereby avowing it, might it push me instead toward reporting that belief in words such as “I **(p.90)** apparently believe that Mogli is honest,” “I’m pretty sure that I believe in Mogli’s honesty,” or “My belief seems to be that Mogli is honest?”³

It may appear that playing safe in this way would be attractive. In reporting the ground-level belief that Mogli is honest, I would be expressing a higher-order belief to the effect that I have that belief; expressive or pragmatic communication at some level is inevitable. But the cost of my miscommunicating a higher-order belief without being able to invoke a misleading-mind excuse—the cost of miscommunicating a belief about whether I believe that Mogli is honest—would not be very high and would not threaten a heavy reputational loss. It would certainly be much lower than the cost of miscommunicating a belief about the world—a belief about Mogli’s honesty—without being able to invoke a misleading-mind excuse. You and others are liable to rely quite heavily on the truth of any such message about how the world lies according to my beliefs.

Notwithstanding this consideration, the resort to higher-order reports about our attitudes of belief about the world would still be unlikely to attract us. For the very fact that the pragmatic or expressive mode of communication allows of only one excuse for error provides a motive for why any one of us should positively cherish it. By communicating a belief in this manner, as emphasized, it is manifest that I take on a greater risk than if I had reported it: I expose myself to the cost of not being able to explain a miscommunication about that belief by recourse to the allegedly misleading character of my mind. And by manifestly taking on such a risk—by opting voluntarily not just to report the attitude of belief—I make my words more expensive and give you and others firmer ground for expecting them to be true. Why, you may think, would I voluntarily take on that risk unless I was pretty sure **(p.91)** that I would not have to pay the cost of being unable to excuse a possible miscommunication by appeal to a misleading mind?⁴

Suppose you want to know about my belief about Mogli’s honesty. And assume that I am anxious to be able to get you to accept whatever communication I make; I am anxious to be treated as someone whose words are credible, both in

this instance and more generally. If I hedge and say that it seems to me that I believe that he is honest, then it will be clear to you that even if I prove not to have that belief, I will be able to get off the hook—and indeed it will be very easy to get off the hook—by saying that I must have gotten my belief wrong. But in that case, it will be clear that my words are pretty cheap and so not very credible. If I refuse to hedge in that reportive manner, however, and say simply that Mogli is honest, then it will be clear that I have foreclosed access to that easy excuse; that I am taking a considerable risk in communicating my belief in that expressive mode; and that my words therefore are more credible than they would otherwise have been.

It is bound to be appealing for each of us in Erewhon to give the words we utter as much credibility as we can, assuming we are pretty confident about what we say. And that means that communicating pragmatically rather than semantically that we believe something is bound to be an attractive option. Or at least that is bound to be attractive with beliefs that are relevant to our relationships with others, bearing on what they may expect of us in interaction. Taking on the risk of pragmatically communicating such a belief should help to ensure that others actually believe what we say, which is bound to appeal on a number of counts. It is likely to get others to rely on us in the instance in question, which may be important for our other purposes. And it will enable us to **(p.92)** prove reliable in living up to those words, thereby improving our general reputational stock.

To choose the pragmatic or expressive communication of a belief, voluntarily foreclosing the misleading-mind excuse, is to avow that belief by the terminology adopted here. It is to opt for that mode of communication over the salient and acceptable alternative of just reporting it. And presumptively, it is to opt for that form of conveying the belief, at least in part, because it represents a more expensive and so more credible form of communication. Absent this difference, there would be little reason for me or anyone else to opt for avowal; it is the feature that makes an avowal more attractive than a report.

The attraction of avowal in this regard will decrease to the extent that it is an avowal only of a belief in a probabilistically or otherwise qualified content, such as that Mogli is probably honest.⁵ The expense associated with the avowal of an unqualified belief that Mogli is honest turns on the fact that if I do not act as if the belief is true, it will be relatively clear that I do not act as if it is true and that I am exposed to the cost of not being able to invoke the misleading-mind excuse. The avowal is expensive, in other words, just to the extent that any failure to live up to it is testable. And so, it is only to that extent that the avowal is more credible than a report of the presence of the belief.⁶

The avowal of a belief that Mogli is probably honest is not expensive in the same degree as an avowal of the corresponding, unqualified belief, because I might fail to act as if he is probably honest without that failure being detectable. It will

be clear that I fail in that way only if I refuse to rely on Mogli's honesty across a wide range of scenarios and it is going to be difficult for anyone to survey my behavior in different scenarios. **(p.93)** Thus, the avowal is only marginally more expensive than a report that I believe that he is probably honest. And so, it is only marginally more effective in raising the credibility of the words I utter. For this reason, the focus in what follows will be on the avowal of unqualified beliefs, and so on avowals of belief that are relatively expensive: that is, on avowals of beliefs such that it is more or less easy to test for whether someone fails to live up to them.

Not all expressions of belief count as avowals, it should be noted, because not all expressions are suitably voluntary. In reporting rather than expressing the ground-level attitude of belief that *p*, as noticed earlier, I give expression to the higher-order belief that I believe that *p*. Do I avow that higher-order belief, then, in virtue of communicating it expressively? No, I do not.

I will count as avowing it only if, having opted to report that I believe that *p*, I can be taken to exercise an unconstrained choice in choosing to express the belief that I hold that ground-level or 0-order belief rather than reporting that 1-order belief as well. But to report the 1-order belief would be to express the 2-order belief that I hold that belief in turn, thereby opening up an indefinite regress. It won't be clear at any level or order, *n*, why I shouldn't opt for reporting the relevant belief at that level, expressing thereby a belief at level *n*+1. And the prospect of that regress may be taken to make the alternative unacceptable. In reporting that I believe that *p*—that I hold the 0-order belief—it is true that I express the 1-order belief that I hold that belief. But the alternative of reporting that 1-order belief in turn is unacceptable—here be tigers—so I do not choose voluntarily to express it and the expression does not count as an avowal.⁷

(p.94) Like any expression of a belief, voluntary or otherwise, an avowal will be more credible for being manifestly more expensive than a report. But the fact that it is a voluntary expression of the belief, adopted on the very grounds of being manifestly more expensive, will make it more credible still. It will show that in taking on that expense I am backing myself not to have to suffer the cost of being unable, as a result of the avowal, to excuse a failure to act on the belief. I bet on myself, as a matter of common awareness, to live up to my words. I put my money where my mouth is, giving you the firmest grounds for taking my words at face value. I commit myself, in the game-theory sense of commitment, to holding by the belief at issue.

The confidence required for avowing beliefs

These observations suggest that given the ready availability of avowing as distinct from just reporting our beliefs, we in Erewhon are likely to cherish the possibility of avowal; we are likely to rely on avowal to give as much credibility

to our communications about our beliefs as our confidence allows. This is certainly going to be so with communications where the belief communicated is likely to matter in continuing interaction with others; this might not be true, to pick a random example, with the belief that I once saw clouds take the form of a mathematical equation. But how can we ever be confident enough to put aside the possibility of invoking a misleading-mind excuse for a failure to display a belief avowed? It is one thing for us to have a means of avowing our beliefs, and a motive for doing so. It is quite another for us to be in a position where we have sufficient confidence about what we believe to be able sensibly to take this line.

What might make it possible for me, then, to have the level of confidence required for an avowal that I will continue to hold this or that belief? What might enable me to have sufficient confidence that I believe **(p.95)** that Mogli is honest, for example, or that the berries on the hill are ripening, or that the weather up north is improving? With any such belief, it is possible that despite appearances I do not actually hold it; holding it is not tied up with who I am, for example, in the manner of religious beliefs or framework beliefs (more on these later). So, what might make it possible for me to be sure enough of holding the belief that I am prepared to put aside the misleading-mind excuse?

This question is challenging, since it may seem that the confidence required could only be available on the basis of a non-naturalistic, Cartesian story about how we know our minds. But it turns out that there is a much more plausible, non-Cartesian answer and that this does not offend in any way against naturalistic assumptions. In order to present this answer, however, it is necessary to turn briefly to more general considerations about language and about what language makes possible.

In order to serve a reporting function in any community, natural language must provide the means for speakers like you and me to communicate how things are, according to our beliefs. And this means that when we take care to determine whether or not it is the case that *p*, to pick an arbitrary sentence from our language, and then assent to the proposition that *p*, we must tend in general to hold the belief that *p*.

If this were not generally the case, then our conscientious, sincere assertions—our careful, truthful reports about the world—would be correlated only contingently randomly with our beliefs and with their evidential inputs and behavioral outputs. And in that case the reports would not serve a useful function in guiding others about what to expect of us. Indeed, the assertions would not even be interpretable. They would not be guided by reliable connections with the prompting conditions that lead me to form beliefs or with the prompted actions that my beliefs occasion; and floating free of such

constraints, it would be difficult for others to make sense of them (Davidson 1984).

What is it going to mean, whether in Erewhon or elsewhere, for me to take care about determining that it is or is not the case that *p*? It can only mean taking evidential care: registering and responding to the data on whether or not *p*, where these are mediated by perception or memory, existing beliefs or the testimony of others. In particular, it **(p.96)** means guarding where needed against distractions that might lead me not to register all the data relevant or not to attend to them in the full manner required for responding as they can dispose me to respond. If I register and respond to the data in that manner, I will be led to assent to or dissent from the proposition before me, or to withhold judgment. And in that case, presumably, the corresponding belief-state will materialize within me.

Thus, if I find myself deferring to the data and assenting to the proposition that *p* after taking such evidential care, then I will presumably come to form the belief that *p*, whether for the first time or in reaffirmation of a belief already held. In either case I will now believe that *p* in the sense, roughly, of being disposed to act and adjust as if it were the case that *p*. Across scenarios where my evidence for '*p*' remains unchanged—and just across those scenarios—I will be disposed more or less dependably or robustly to take actions and draw inferences as if it were the case that *p*.⁸

More specifically, I will be disposed to do this robustly over a range of variations in the distractions that may lead to oversight or inattention. Evidential care is designed, where needed, to guard against such distractors. If I did not form a suitably robust or dependable disposition as a result of carefully assenting to '*p*', then the words I used in making the corresponding report would not be indicative of my attitude or even intelligibly interpretable by others.

There is no problem about how I and others might acquire the concept of a distractor in Erewhon. For the criteria by which something counts as a distractor will be available to each of us just on the basis of understanding our practices of assertion. The role that data or evidence **(p.97)** play in relation to reports about the world presupposes that distractions are possible and that reports are guided properly by data only in their absence. That fact is bound to be accessible to me and others in Erewhon insofar as we make reports to one another. And so, the concept of a distractor, with the associated ideas of oversight and inattention, will also be accessible.

The upshot of these general considerations appears to be that I can be very confident that I believe something when I find that I assent to it after taking any evidential care needed over the relevant data. That conclusion is independently plausible, being borne out by the fact that people often answer the question as

to whether or not they believe that *p* by thinking about whether it is the case that *p*—not whether they believe that *p*—and answering, when appropriate: “*p*.” In other words, they routinely treat it as a question about whether the data elicit a belief that *p* in them, not whether introspective or other evidence reveals the presence of the belief (Evans 1982; Byrne 2011).

But still, the conclusion needs to be qualified, for there is one important complexity to be added to the account of how I can be confident enough about whether or not I believe something to be ready to avow it. This, to anticipate, is that I not only seek to overcome potential distractors, paying attention to all the relevant evidence in forming a belief that *p*; I also try to deal with what may be described as potential disturbers of my performance as someone who avows the belief that *p*.

The possibility of disturbance arises with at least some beliefs that matter in continuing interaction with others. To illustrate the possibility, suppose I am brought by careful consideration of the data to assent to the proposition that the gambler’s fallacy is a fallacy. It may be, first, that the belief is liable to disappear in the excitement of the casino, assuming that there is gambling in Erewhon: when there is a run of blacks, I am likely to commit the fallacy, growing in confidence that a red will come up next. Yet, second, I would certainly not be disposed in the wake of such a performance to invoke the excuse that I temporarily changed my mind about the matter during my visit to the casino. On the contrary, I would continue to maintain that the fallacy is indeed a **(p.98)** fallacy and disown my behavior in the casino; treat it as a failure to respond to the data that show the fallacy to be a fallacy.⁹

Disturbers of belief, in the sense illustrated, are characterized by three features. First, like the excitement of the casino, they are liable to cause a person not to act on a belief they avow. Second, they do not lead the person to excuse the failure to act appropriately by appeal to a temporary change of mind. And third, there are various executive steps that anyone can take in order to neutralize or avoid their impact; the influence they exercise is not inescapable. Thus, to illustrate this third feature, I can guard against the effect of being in the casino by bringing along a friend to remind me of the fallacy at appropriate points; or, of course, I can avoid casinos altogether.

As in the case of distractors, the criteria by which something counts as a disturber are internal to my practice in assertion. The notion of an assertion or report that is maintained over time, and its connection with the idea of an enduring belief, highlights the possibility of disturbance, and a consequent failure in assertoric practice. That fact is bound to be available to me and others in Erewhon insofar as we make a habit of giving and being given reports.

Disturbers constitute a potential danger with any beliefs that—like the belief that the gambler’s fallacy is a fallacy—may be expected to continue over time and to matter in my relationships with others. They may take a variety of forms, ranging from the effects of wishful thinking and gullibility to those of incaution and prejudice and are well illustrated by many of the biases and heuristics that are known to shape human cognitive performance (Gilovich, Griffin and Kahneman 2002).

If I do nothing to guard against it, the prospect of failing in the casino means that I may not be sufficiently confident to avow the belief that the fallacy is a fallacy, inviting others to rely on my displaying it. And so, a more nuanced story is needed for the confidence required for avowing a belief with such a role in relations with others. As someone **(p.99)** competent in the natural language of Erewhon, I can be confident that I believe that p in a suitably reliable way just when I satisfy two conditions. First, I am evidentially careful, in the measure needed, to register and respond to relevant data—to avoid distractions—before I assent to it. And second, I am executively careful, in the measure needed, to maintain my sensitivity to the data: that is, to avoid disturbers.

Suppose I satisfy these two conditions, that the data continue to support a belief that p, and that I continue to be a competent subject. In that case I will sustain the belief robustly over variations in factors other than the support provided by the data, and the competence I display in responding to them. In particular, I will sustain it robustly over variations in the distractions and disturbances that might come into play and affect me.¹⁰

How can I muster sufficient confidence, then, to be able to avow a belief that p as distinct from merely reporting that I believe it? I must take care, as needed, to register all relevant data without distraction and to nullify potential disturbance—in short, I must take any form of evidential or executive care that is required—and I must find that the data I thereby register elicit my assent to the proposition that p. Evidential and executive care will not be required in many cases, of course. With common beliefs, whether about the character of my neighbors, the layout of my environment or wholly abstract issues, I will not have to worry much about being distracted from relevant data, or about being disturbed in my sensitivity to them. I will take special pains to exercise **(p.100)** evidential and executive care only when there are independent reasons to be concerned.

Assume that I defer with due care to data that support a claim that p, taking whatever precautions are required against distraction or disturbance. And assume that I can know what I am doing in taking such an action, whether it be described as assenting to the proposition, or making up my mind about it.¹¹ That knowledge is going to be enough to support a high degree of confidence, arguably indeed a claim to know, that I believe that p. Having made up my mind,

I can claim, in a seventeenth-century phrase, to have a maker's knowledge of believing that *p*, not the knowledge of an observer, even an introspective observer.¹² I can speak for what I believe with an authority of a special, practical sort (McGeer 1996; Moran 1997, 2001; McGeer 2008).

If I have the means of avowing a belief that *p*, and a motive for doing so, then these observations suggest that I will often go ahead and avow it; I will not be inhibited by a lack of confidence. Avowing the belief will make sense whenever there are no grounds for worrying that I may not be exercising sufficient evidential or executive care. Either I do take pains to exercise such care, or it is clear that no special care is required.

Consistently with this picture, we in Erewhon may not always perform to par. Short of having suitable grounds available for avowing a belief that *p*, for example, I may be pushed into avowing it by distractions that generate oversight or inattention. Or I might be led to form the belief by disturbing factors like wishful thinking and prejudice; these may affect the formation of beliefs, as well as disturbing beliefs **(p.101)** already formed. And if I succumb, I may also manage to persuade myself that suitable grounds for avowal are actually available.

Nothing in this narrative precludes these sorts of self-deceptive possibilities, and nothing, as will appear, precludes parallel possibilities in the case of avowing a desire or intention. All that the narrative assumes is that the possibilities will not be rampant, driving out avowals of the bona fide sort. And that assumption is surely plausible.

3.2 The avowal of desire and intention

Desires and intentions

As it is going to be manifestly attractive for me and others in Erewhon to avow our beliefs rather than just report them, so the same is true for the other attitudes we may wish to communicate to one another. Or at least that will be so in the case of attitudes that are important in our relationships with others and that we must want others to recognize in us. The focus in this discussion will be on desires and intentions, but the line taken can be extended to other attitudes as well, such as those, for example, of hope and affection and trust.¹³

There is a difference between desire and intention, as those terms will be employed here. An intention is a final disposition to choose one of a given set of alternatives in a choice, whether the choice be between different act-options, different routines for making decisions, different principles for constraining decisions, or whatever.¹⁴ That it is a final **(p.102)** disposition to choose this or that option will show up in the fact that it makes a fixture of that choice in the further deliberations of the agent; short of second thoughts, the agent may be

expected, where relevant, to deliberate on the assumption that the chosen option is going to be in place (Bratman 1987, 1999).¹⁵

A desire is also a disposition to choose, if a choice is available, but it is not restricted in the objects that it may target. It may be a desire that the world in general be of such and such a character, a desire that other people act in a certain manner, or indeed a desire to pick one or another option in a personal choice. Such a desire will approximate an intention if it assumes the form of a final disposition to choose that option.¹⁶

The attraction of avowal

Suppose, then, that I am confident that I have a wish to prove reliable to others, a preference that people should talk about differences rather than squabbling over them, or an intention or plan to go hunting tomorrow. Such desires and intentions are likely to be of importance in my interacting with others. They are different in that respect from, for example, a desire I experienced on some particular occasion to fly like a bird or from a policy of reading novels when alone.

(p.103) If I want to convey the relevant sort of desire or intention to you with a suitable degree of credibility, it will be useful to be able to avow the attitude rather than just report it. Avowing it will mean communicating that I have it in such a manner that I cannot excuse a failure to live up to it by claiming that I must have been misled about my attitude. And such a mode of communication will give you much firmer ground for taking me at my word, relying on my possession of the attitude, than if I reported on its presence in a way that kept that excuse open.

At least that will be so the extent that a failure to act on the desire or intention avowed is testable. As noted earlier, avowing a probabilistically qualified belief is not going to be much more expensive than reporting it, since it is difficult to test for the presence of such a belief, and is not going to make the communication much more credible. And by the same reasoning, avowing a desire or intention is not going to be much more expensive or credible than reporting it, if it is relatively difficult to test for a failure to act on that desire or intention.

This observation explains why the focus in the evolving narrative is on the avowal of a restricted class of desires and intentions, as the focus is on the avowal of unqualified rather than qualified beliefs. In the case of desire, the focus will be on avowals of desires that connect closely with scenarios for action that the avowing agent is likely to face, not on idle wishes and the like; while the agent may not always act on such a desire, there will normally be a salient consideration in place to explain the failure. And, in the case of both desire and intention, the focus will be on avowals of unconditional desires and intentions, like the desire or intention to do something, X, not on conditional counterparts:

not, for example, on the desire or intention, if p or r or s, to X; it will be possible to test for the presence of such attitudes, only when the antecedent happens to be fulfilled.

Given this focus, it is clear why the avowal of desires and intentions is going to be an attractive option for us in Erewhon. It will have the same appeal as the avowal of beliefs, or at least of probabilistically unqualified beliefs. But there is a problem about how we in Erewhon can find the means of avowing desires or intentions and, even if we do find a means of doing so, about how we can have the high level of confidence about those attitudes that is required for avowal.

(p.104) *The means of avowing desires and intentions*

I can avow a belief that p, as we know, by asserting that p, thereby expressing my belief-state. But I cannot avow a desire or intention that q by asserting that q; such an assertion would express a belief that q rather than a desire or intention that q. While I may have a strong motive for avowing desires and intentions, then, I may lack the means of doing so. Certainly, I cannot avow a desire or intention that something be the case—say, that it be the case that q—by using a sentence like “q” to assert that that is the case.¹⁷

Notwithstanding this difficulty, however, it is possible to discern a relatively straightforward means whereby we would be able to avow desires and intentions. In order to make this means visible, however, it is necessary to return briefly to the situation with the avowal of beliefs.

The expressive means of avowing a belief is of fundamental importance because it is going to be saliently available as well as saliently attractive in Erewhon, even at the stage where we are exclusively interested in making worldly reports to one another. But suppose that the expressive avowal of belief has become standard practice in Erewhon, as the preceding argument suggests that it would become standard practice. Suppose that it has become a matter of common awareness, in other words, that in Erewhon we will generally want to avow the beliefs we hold—or at least the beliefs in which others have an interest—and that a standard way of doing this is just to express those beliefs: to say “p” in communicating that we believe that p.

Under those circumstances, it is plausible that we will begin to recognize other means of avowing our beliefs. And it turns out that those other means of avowing beliefs offer us models for the avowal of desires and intentions as well.

The attraction of avowing beliefs that are important in our relationships with others is going to be obvious to all of us in Erewhon. And so, it is likely to be a matter of common awareness that it has this appeal; **(p.105)** after all, we will each have access to evidence of that appeal, to evidence that we each have access to that evidence, and so on. But if this is a matter of common awareness, then the default assumption we will each make with others is that in

communicating relevant beliefs to us they are meaning to avow them. They are meaning to speak for what they believe while putting aside the possibility that they may have been misled about their own minds and gotten those beliefs wrong.

Suppose then that in the presence of that default assumption, I do not say “p” in expressive mode but resort to the ascriptive mode, as in saying “I believe that p.” Should I be taken to be merely reporting on my belief rather than avowing it? In many contexts, especially contexts in which I act as if I am willing to avow the belief, it will make sense for you to treat my self-ascription of the belief as if it had the force of an expression, not a report: as if it amounted to an avowal. In those contexts, you would expect me to go out of my way to indicate that I was merely reporting on the belief, if that was indeed my intention. You would expect me to resort to cautious phrasing, as in saying that it seems to me that I believe that p, or something of the kind.

Absent such phrasing, you will naturally take an ascriptive assertion like “I believe that p”—an assertion in which I ascribe the belief to myself—to have the same avowal force as the expressive assertion “p.” And equally you are likely to assign the force of an avowal to other remarks, too: say, to an explanatory remark such as “The data explain why it is the case that p.” In either sort of case, ascriptive or explanatory, you will expect me to be ready to stand by the belief, and not to hedge in the manner of a self-reporter. Hedging in that manner would be unusual enough for you to expect that I would do more to indicate that I was hedging, if indeed that is what I was wanting to do.

That this is what would happen in Erewhon is borne out by the fact that this is what happens in actual languages. As in actual usage, I could hardly expect to be taken as a mere self-reporter if I said that I believed that Mogli was honest. In order to mark out my utterance as merely reporting the presence of the belief, I would have to say that my own impression is that he is honest, or that I am inclined to credit him with honesty, or something of that markedly cautious kind.

(p.106) Assuming that this line of argument is sound, consider now the point at which we in Erewhon have established a practice that allows us to avow our beliefs in ascriptive and explanatory assertions as well as in expressive ones. At that point, plausibly, we will also have provided ourselves with a salient means of avowing desires and intentions.

Saying in ascriptive mode “I desire to be reliable”—or, equivalently, “Being reliable attracts me”—is not necessarily going to be taken as a mere report that I have that desire but may be heard in appropriate contexts as an avowal. And the same will be true of saying in similar mode that I prefer talking to squabbling, or that I intend to go on a hunt tomorrow. Again, saying that there

are factors that explain why I hold such a desire or intention is not necessarily going to count as a detached explanation but may be taken in suitable contexts as an avowal of the attitude explained.¹⁸

Or at least this will be so with communications in which the attitudes I convey are likely to be of importance in my interactions with others, unlike the episodic desire to fly or the policy of reading novels when alone. Thus, I will be expected in important cases to go out of my way to indicate that I am hedging my bets if that is what I am doing in communicating the attitude. I will be expected to resort to quaint phrasings, as in saying, "Looking at my behavior, I seem to prefer talking to squabbling," ' or "I'm sure that my intention is to go to the hunt tomorrow," or, of course, "I'll probably go on the hunt tomorrow."¹⁹

(p.107) *The confidence required for avowing desires and intentions*

But it is one thing to show that, like others, I will have a motive and a means of avowing desires and intentions in Erehon. It is quite another to show that I can be confident enough of having such an attitude to be willing to avow it: to be willing to discount the possibility that I am misled about my own disposition. In the case of belief, I can find a sufficient basis for confidence, absent distraction or disturbance, in the fact that the data to which I defer with suitable care elicit assent to the proposition. In order to avow a desire or intention, I need similar grounds to be confident about holding it. But where might I find an effective basis for confidence in this case?

The most plausible answer, which fits with a long tradition of thinking, is that I can find such a basis in the sorts of properties that typically elicit desire or intention—for short, in the desiderata or attractors present. Thus, I can be generally sure that I desire or intend R insofar as the properties that attract me to it here and now, assuming I have taken care over registering and responding to these, dispose me to act and adjust dependably as if R were suitably attractive. As already registered, I can be generally sure of believing that p insofar as the data dispose me to respond dependably as if it were the case that p. And I can be generally sure of desiring or intending R insofar as a parallel condition is **(p.108)** satisfied: the desiderata, carefully rehearsed, dispose me to respond dependably as if R were attractive.²⁰

This claim holds true of desire and intention, of course, as the other claim holds true in the case of belief, only in the absence of distraction and disturbance. Thus, in this case, as in the case of belief, I must take evidential care against the distraction that would prompt oversight or inattention in relation to desiderata, and executive care against the disturbance of the desire or intention they lead me to form. Or at least I must have reason to think that no particular care is required.

Disturbers of desire and intention must satisfy three conditions. Their presence can prompt a failure to display the desire or intention communicated; they do not lead me to explain the miscommunication as a temporary change of mind; and I can do something to guard against them.²¹ The possibility of disturbance in the case of a belief was illustrated with the illusion that may take hold of me in the casino, undermining the effect of my assenting to the claim that the gambler's fallacy is a fallacy. The possibility of disturbance in the case of desire or intention is even more salient, since psychological factors like anxiety or impulse or sheer laziness provide ready examples.

As the confidence needed for the avowal of belief requires evidential care against distractions and executive care against such disturbers, so the same is true here. I must take precautions, where needed, to register without distraction the desiderata on which I rely and to guard myself against potential disturbers.

(p.109) *The nature and role of desiderata*

The desiderata that serve to elicit desire or intention come in many different forms. They include neutral properties that can make a scenario attractive for anyone in any situation: that it would be fun, that it would secure peace, that it would reduce suffering. They include agent-relative properties that can make a prospect attractive for anyone in a certain relationship or position: that it would create an advantage for my child, for example, or further the prosperity of my tribal group. And they even include agent-relative properties that make things attractive, conditionally on the presence of a contingent need or taste: the property of constituting a drink in the presence of thirst, for example, or of giving me relief in the presence of pain.²²

All of these desiderata, even those that presuppose a need or taste, reflect properties of the prospects targeted in desire and intention. They are properties that elicit my desire or intention and that dispose me to act and adjust more or less dependably as if the prospects were attractive. Assuming that those desiderata remain in place, and I continue to be a competent subject, I will be disposed in the appropriate way—I will hold the relevant desire or intention—robustly over variation in other factors; in particular, variation in the distractions or disturbances that might affect me.

To hold that desires and intentions can be grounded in desiderata, as beliefs can be grounded in data, is to go along with the idea, long accepted in philosophical tradition, that there are general considerations that lead human beings to form desires or intentions and that agents consciously or unconsciously let such considerations determine their **(p.110)** choices. The attitudes do not appear out of the blue, so this orthodoxy holds, but are generally elicited by attractive features that people identify in the targets of their desires or intentions: features that they agree in finding attractive, even if they weigh them differently against one another. If someone seeks a particular outcome, so the idea goes, they will

always do so, at least absent distraction or disturbance, in explicit or implicit response to the attractive features it promises to realize.

This orthodox view does not hold that it is just a brute empirical fact that human beings are moved to seek one or another prospect by the general desiderata that it promises to realize. The idea is that human beings would not be able to make proper sense of the desires and intentions they ascribe to others, or even to themselves, unless there were desiderata in view—perhaps desiderata weighed in different ways by different people—that explained the presence of those attitudes. They would fail to make sense of the ascriptions in the way in which they would surely fail to make sense of the ascriptions of certain beliefs—empirically vulnerable beliefs—if there were no data in view that explained why the beliefs should be maintained. Thus, they would find the person who self-ascribed attitudes of a data-insensitive or desiderata-insensitive kind more or less unconvertible—incapable, to put the idea roughly, of being engaged in mutually profitable exchange (Pettit and Smith 1996).

Elizabeth Anscombe (1957) uses an eye-catching example to argue that intelligible attitudes of desire or intention—that is, attitudes whose ascription to someone would be intelligible—have to be grounded in familiar desiderata. She asks how others would think of a person who seeks something as unlikely as a saucer of mud but cannot do anything to make sense of that desire, presenting it to them under an aspect with recognizable attractor potential. In order to find the person conversable, she suggests, others would have to see some general aspect under which the saucer of mud appeals: say, as an ornament or as a reminder of mortality. They would not themselves have to be moved by the prospect of possessing that particular ornament or reminder, but they would have to be able to see why the property of being an ornament or a reminder **(p. 111)** of mortality might have an appeal: why it might have the capacity to make something attractive.²³

These observations suggest that when desiderata are invoked to explain why I should desire or intend something, they present me as an instance of a familiar, universal type. If a preference is explicable by such desiderata, then that means that I do not seek the end at issue because of its pure particularity; I seek it because of the type of end in view, or because of the type of relationship I bear to that end. Thus, I seek to reduce someone's suffering because of what the reduction of anyone's suffering involves or, if it matters that the person is a friend, because of what the reduction of that friend's suffering involves. The explanation makes me intelligible to others, then, by presenting me as an instance of a familiar type: I am the kind of person concerned about suffering or the kind of person concerned about the suffering of a friend.²⁴

The picture according to which desires and intentions relate to desiderata as beliefs relate to data is not endorsed on all sides. Decision theorists reject the picture insofar as they treat preferences as primitive rankings, ignoring the possibility that reliable attractors or desiderata lie at their origin. But they reject it, arguably, only by way of a convenient simplification, not as a position that is objectionable on independent grounds (Pettit 1991b; Dietrich and List 2013).

Opponents of the picture also include particularists, as they are often known (Dancy 2004). While they agree that the properties of objects of desire or intention may play a characteristic role in eliciting that attitude—this, by extrapolation from the role they are supposed to play in eliciting corresponding moral judgments—they deny that those properties always weigh in the same direction. Thus, in a much-cited example, the pleasure of an innocent activity may weigh in its favor, **(p.112)** the pleasure of doing something noxious, like torturing another, may weigh against.

Particularism of this kind is counterintuitive, since it has to reject the common-sense idea that in deliberation about choices, we human beings explicitly or implicitly weigh the pros and cons attaching to each option and form a desire or intention on the basis of the resultant effect. The main arguments in its favor draw on examples such as the pleasure case. But those examples need not be taken to undermine our standard view of deliberation and argue for particularism. On a rival interpretation, they demonstrate that it is not pleasure as such that counts as a desideratum with us but, rather, innocent pleasure.²⁵

Back to the confidence issue

It is time to return, finally, to the issue of confidence. As I can form a belief by deferring to relevant data, so I can form a desire or intention insofar as I defer to corresponding desiderata. Thus, I can have confidence that I hold the belief in question, whether for the first time or not, by virtue of knowing that I am deferring to the data and am on guard against distraction and disturbance. And I can have confidence that I hold the desire or intention, whether for the first time or not, by virtue of knowing that I am deferring to the desiderata and am on guard in a parallel way against distraction and disturbance. In this case, as in the other, the confidence needed for avowal materializes by virtue of having a maker's knowledge of the attitude I hold.

The observations rehearsed identify grounds sufficient for me to have confidence enough about desiring R or intending to X to be ready **(p.113)** to avow that attitude. But, as in the belief case, it is important to recognize that things may not always match this ideal. Thus, I might avow such a desire or intention, and even think I had grounds for doing so, without actually having access to such grounds. The attitude might be one that I embraced, not because of the attraction of the object, but rather because of the sort of oversight or inattention

that distraction can produce, or because of the disturbing influence of impulse or whim or the like.

As in the belief case, nothing in the narrative precludes these sorts of self-deceptive possibilities. All that the narrative assumes is that in Erewhon the possibilities will not swamp the bona fide cases. And that assumption is as plausible in this case as in the other.

3.3 The pledging of intention

The idea of pledging

By the definitions given earlier, to make a pledge as distinct from an avowal in communicating an attitude is to go one stage further in reducing possible ways of excusing a miscommunication. It voluntarily and manifestly forecloses not just the possibility of excusing a failure to live up to it by reference to a misleading mind but also the possibility of doing so by reference to a changed mind. If I avow the intention of going with you on a hunt, as in an earlier example, then I can scarcely excuse my failure to turn up by saying that I was misled about my intention. But in that case I can certainly excuse it by saying that I changed my mind since speaking to you. If I pledge the intention to join you on the hunt, however, then I cannot avail myself of this excuse, either.

It should be clear that we are each going to have a motive in Erewhon for pledging attitudes to one another, if pledging is indeed possible. In particular, we are going to have a motive for pledging the congenial or collaborative attitudes that matter in building or maintaining relationships with one another. Pledging an attitude is even more expensive than avowing it, since it exposes me to a greater risk of not having any excuse for failing to act on the attitude. The **(p. 114)** pledge will be highly credible because of the risk that I choose to take in making it. And it will be all the more credible because of the fact that in opting for it voluntarily and manifestly, I convey the message that I fully recognize the cost of failure but nevertheless back myself not to incur it; I commit myself to being someone who lives up to the attitude I convey.

As it will clearly be attractive in many cases to pledge our attitudes to one another in Erewhon, so it should be clear that if we are bent on doing so, then we will be able to find a linguistic means of signaling that we are making a pledge and not an avowal or report. The fact that, absent expressive possibilities, I can avow a desire or intention in an ascriptive or explanatory fashion shows just how flexible and multifunctional our utterances can become. It argues plausibly that in the presence of an assumption that it is possible and attractive for me to pledge an attitude rather than just avow it, there will be forms of utterance I can invoke to communicate that I am indeed making a pledge.

But even if there is no problem about how we may have a motive for pledging any attitudes and may be able to construct a suitable means of doing so, there is still a question as to whether we could ever garner enough confidence about our attitudes to be able to pledge them. It turns out that while we may be able to have confidence enough to pledge an intention, we could never have enough confidence to be able to pledge other attitudes.

The problem with pledging a belief or desire

Might I be able to pledge a belief? In particular, might I be able to pledge a regular, empirically vulnerable belief of the kind that I recognize I might not maintain? We live in a changing, incompletely grasped world and although I may think that the data are sufficient to elicit belief in an empirical proposition 'p', enabling me to avow a belief in it, I can never be sure that the data will not later be overturned or outweighed. Indeed, for me to consider pledging such a belief would betray a misconception about the very attitude of belief. It would show that I did not treat it as responsive to potentially changing data.

(p.115) As against this, however, aren't we in Erewhon likely, in a fashion familiar from the actual world, to pledge religious or political beliefs or even beliefs that define the framework in which we think, such as beliefs in principles of classical logic like those of non-contradiction and excluded middle? Yes, but the best gloss on such a pledge is that it involves pledging an intention: say, the intention to treat certain texts or authorities or frameworks as definitive, letting them shape the construal to be given to any other sources of evidence. In particular, it involves pledging the intention of continuing to take that line, conditionally on the absence of a radical conversion or disruption in our way of thinking; while we may think that this is extremely unlikely, we can hardly discount the possibility altogether.

Might I be able to pledge an attitude, not of belief, but of desire? In order to do so, I would have to be able to identify relevant desiderata or attractors in the object of desire. And in deferring to those desiderata, I would have to be confident enough about their remaining effective—and about my ability to guard against distraction or disturbance—to be able to foreclose the changed-mind excuse as well as the misleading-mind excuse. Is there any desideratum that might help me to muster such confidence?

Surprisingly, there is. If I pledged a desire, then the very fact of making the pledge would bring a desideratum or attractor into existence that might serve in the required role. It would make it the case that sticking with the attitude has at least this appealing feature: that it would show that I can be relied upon to keep my word. The question, then, is whether I could rely on that feature to enable me to pledge a desire for hunting, for example; that is, an attraction to hunting that is based on the pleasure it provides.

The answer is that I could not. Suppose that I pledge a desire for hunting; that many of the desiderata that attracted me to hunting cease to be appealing; but that I continue to choose hunting because of wanting to show that my word is my bond. Would the desire for hunting remain in place as a result of the pledge?

No, it would not. I cannot count as desiring to hunt in the relevant sense when I only continue to choose it because of having given my **(p.116)** word. Desire in the sense at issue requires me to be attached to hunting on the basis of desiderata other than the attractor that a pledge would put in place. It requires me to like hunting for the sport or recreation it offers, for the chance it gives me to get out into the country, or something of that sort. If hunting were to lose those attractions for me, then even if I continued to hunt because of having given my word, I would not continue to desire or like hunting in the relevant sense. And so, I cannot pledge to hold by a desire any more than I can pledge to hold by a belief.

The possibility of pledging an intention

This problem does not arise, however, with an intention or plan or anything of that kind. Suppose that in speaking with you I pledge an intention or plan to join you on the hunt—equivalently, I pledge to join you—wanting the thrill of chasing prey over open sunny spaces. And imagine that it rains heavily on the appointed day, but that I turn up nevertheless because of having given you my word. Do I count as still holding and acting on the intention pledged? Yes, I do. With an intention as distinct from a desire or preference, the attitude does not have to be sourced in certain sorts of desiderata in order to count as remaining in place; that I act as the intention requires—presumably because it requires it—is enough to establish the presence of the intention.²⁶ And so, the attractor that pledging an attitude creates in favor of maintaining the attitude can serve in this case—although only, it appears, in this case—to give me the confidence required for being able to make a pledge.

The possibility of avowing beliefs and desires was defended earlier on the basis, not of a Cartesian form of self-access, but of a naturalistically intelligible form of maker's knowledge. What now transpires is that the possibility of pledging attitudes, in particular intentions, is defensible **(p.117)** in a parallel way without recourse to a non-naturalistic, libertarian account of self-control. All that needs to be presupposed in order to make sense of how I can pledge an intention is that I can recognize the motivating effect of certain desiderata, in particular the desideratum of proving faithful to my word.

According to previous arguments, I put myself in a position to avow a belief or desire or intention on the basis of consciously deferring to a suitable body of data or set of desiderata, where I am careful to register the data or desiderata without distraction and to guard against the possibility of disturbance. I know that I believe or desire something with sufficient confidence to be able to avow

that attitude by virtue of knowing that I defer to those data or desiderata: this gives me a maker's knowledge of the attitude avowed. The same sort of maker's knowledge will enable me to recognize that I intend something with confidence sufficient to be able to pledge the intention; that is, to pledge to act as the intention requires. In consciously recognizing and deferring to the desideratum that the very act of pledging brings into play—the desideratum that consists in proving to live up to my word—I can achieve the degree of confidence required.

As in the case of avowing, the sort of pledging that will be relevant in the narrative constructed here involves pledging an intention of an unconditional kind—say, an intention to X—not a conditional intention: not an intention, if p, to X; not for example an intention, if the weather is good, to join you on the hunt. The same considerations apply here as with the avowal of desire and intention, and with the avowal of belief. To the extent that the intention I pledge is conditional, it will often be hard to test for whether I fail to display it; testing for it will require the satisfaction of the antecedent. And to the extent that that is so, the pledge will involve less risk and expense, and will boost the credibility in a lesser measure.

It is worth noting that conditional intentions would allow us to make pledges that involve beliefs and desires indirectly. Thus, consistently with the narrative presented, I might pledge an intention, if the evidence is supportive, to believe that p or, if suitable desiderata are in place, to desire R. It is in this sense, to revert to an earlier comment, **(p.118)** that I might pledge to hold by a framework belief, such as a belief in classical logic, treating it as necessary or self-evident. What I pledge in such a case is the intention, if this does not lead to problems of a kind I can barely foresee, to stick with classical principles in organizing my responses to data (Quine 1970).

Pledging versus promising and predicting

The notion of pledging an intention to X—pledging to act on the intention—corresponds to the more regular idea of promising to X. But the notion of promising in ordinary usage has a strong moral or ethical flavor, as remarked earlier. It is represented as an act such that if I make a promise to do something, then I have an ethical obligation, however defeasible, to do it.

Pledging, as introduced at this point, has no such ethical connotations. When I make a pledge in Erewhon, as when I make an avowal, I back myself to act as thereby advertised, manifestly exposing myself to serious reputational costs in the event of failure. What I do is more akin to making a side-bet that I will hold and act on the intention pledged—a side-bet strategically designed to entice you and others to rely on me—than it is to giving you a promise in the ordinary, moralized sense of that term.

As pledging to act on an intention should be distinguished from the moral possibility of promising to act that way, so it should be distinguished from the empirical possibility of predicting the action. It may be that certain intentions or plans, perhaps again of a religious or political or framework kind, are relative fixtures in my psychology such that I can be fairly well assured—if you like, assured in the way an observer might be assured—that they will continue in place and I can predict with confidence that I will behave as they require. But pledging in the sense introduced here is quite distinct from such prediction, as it can materialize with intentions that are not rooted in the same way in my psychology. Indeed, pledging assumes that the intention it supports is not a fixture of that kind; it serves a useful purpose just to the extent **(p.119)** that the attitude pledged is one that I might fail to maintain in the absence of a pledge.²⁷

Pledging an unconditional intention, by the account offered here, is more costly, and hence more credible, than merely avowing the intention. If I pledge to act on a certain intention, as in pledging to join you on the hunt, then my stake in living up to those words is higher than my stake would have been, had I merely avowed an intention to join you. And hence you can rely with greater assurance on my joining you than if I had just made an avowal.

As in the case of avowal, of course, the cost of pledging need not make the act prohibitively expensive. If I fail to join you on the hunt but can invoke the practical, unforfeited excuse of a broken leg in explanation of the failure, then I do not lose my stake. And the same is going to be true when I can plead an exempting disability like a temporary bout of amnesia to explain the failure. Any such factor can persuade you that, despite the failure, you need not despair of me as a cooperative and reliable interlocutor. **(p.120)**

Notes:

(1.) The notions of common awareness and of voluntariness were introduced in Chapter 2.

(2.) This suggests that the standard analysis of communication may be unnecessarily strong. Assuming overtness, all that may be required, as in the communication that I believe what I say, is at the primary level that I convey the information intentionally and, at the secondary, that I intentionally make it salient that I am intentionally doing that. Doing something, Y, intentionally amounts to doing it, roughly, as a means of satisfying my desires according to my beliefs. But doing Y intentionally in that sense does not strictly require having an intention to do it as such. It may only involve having an intention to do X as such, while recognizing that to do X is inevitably do Y (Scanlon 2008, 8). I desire and intentionally realize the X-Y package, but I may do so because of

desiring and intending X and despite not having such an attitude to Y (Pettit 2018a).

(3.) Why report the beliefs in such words? Why not say simply “I believe in Mogli’s honesty” or “My belief is that he is honest?” The reason will become clear later: viz., that such self-ascriptions of the belief are likely to be taken as avowals rather than reports.

(4.) The linkage between expense and honesty in animal signaling, which is a continuing subject of debate, is often expressed in the handicap principle (Zahavi and Zahavi 1999). According to this principle, roughly, the relative expense of a wasteful or handicapping signal—say, the peacock’s tail, the exaggerated jump of the gazelle—conveys the wealth of resources at the animal’s disposal: to the peahen in the first example, to the predator lion in the second. See Maynard Smith and Harper (2004). In such cases the expense is undertaken non-intentionally by nature; in the case of concern to us it is undertaken intentionally by the relevant individual.

(5.) The observation also applies if the avowal is expressed as an avowal of a certain, less than full degree of belief that Mogli is honest.

(6.) This observation has a number of implications that cannot be pursued here. Thus, it may explain why on-off beliefs figure more prominently than credences—degrees of belief—in ordinary discourse. And it may explain why beliefs in religious or broadly transcendental matters, whose presence is not independently testable, are often treated as a matter of personal identity, affiliation or even preference.

(7.) Notice, to return to an observation in Chapter 2, that voluntariness is more demanding than intentionality. Assuming that I report that my 0-order belief that p, it is available as a matter of common awareness that I express and pragmatically communicate the 1-order belief that I believe that p. But while I do not express that 1-order belief voluntarily, for the reasons given in the text, I may do so intentionally. Desiring and intending that I report the 0-order belief that p, I may foresee that reporting that 0-order belief inevitably involves expressing the 1-order belief that I believe that p. And that implies, as we saw earlier, that I may intentionally express that 1-order belief; I may adopt the package deal involved—reporting my 0-level belief and thereby expressing my 1-order belief—as a recognized means of furthering my desires according to my beliefs.

(8.) The assumed concept of belief is broadly functional in character, building on the notion of credence in decision theory; see Stalnaker (1984); Pettit (1998). There is a serious issue as to how credences relate to acts of assent—and to the states of mind that those acts express—but in this context, I ignore the problem. See Pettit (2016b). There is also a serious issue that arises from the fact that

beliefs confront evidence in a network, not one by one; but with the exception of some later comments on framework beliefs, I ignore this as well.

(9.) There are otherwise similar cases, of course, in which I may think of myself as having changed my mind. But in those cases, I will presumably think that certain data—misleading data, as I may judge them later—prompted the change.

(10.) This claim is most readily intelligible on the assumption that the data that support or elicit belief leave only one candidate proposition in place; they do not leave room for me to believe either of a constrained number of alternatives. The assumption that data are determinative in this way, rather than merely constraining, will be maintained for convenience throughout this study. But even if the constraining view were adopted, most of the claims made would survive, albeit in a somewhat modified form. The line taken here means that as I put myself in the position to avow a belief, I put myself equally in a position to provide my interlocutor with assurance that things are as they are represented to be in the belief avowed; on related themes, see Lawlor (2013).

(11.) Thus, it may be that I can know what I am doing, not in virtue of scanning my mind for the guiding intention, but in virtue of being able to counterfactualize reliably about how I would respond to various scenarios. I can know that I am going for a stroll rather than going to the library by knowing that if I found the library closed, that would make no difference to what I did. The ability to counterfactualize in that manner, however it is grounded, is necessary for intentional agency. See Pettit (2016b).

(12.) The seventeenth-century notion of maker's knowledge is prominent in Hobbes and Vico (Pettit 2008a, Ch. 1). Rae Langton (2009, Ch. 13) cites an earlier employment of the idea in Maimonides.

(13.) Avowing that I hope to finish this book by year's end, or that I trust someone, or that I love you may involve the avowal, and even the pledging, of a corresponding intention: to act as if I believe I will finish the book, to act as if I fully rely on the person, to act toward you as love requires. And in such cases, part of the motivation may well be that doing this is likely to have a positive effect on my finishing the book, on your proving reliable, or on our maintaining mutual affection. But such complexities are put aside in this work. For related studies, see Pettit (1995, 2004) and McGeer and Pettit (2017).

(14.) The alternatives that I neglect in favor of intending to do something, X, need only be apparent alternatives. Thus, to return to the Frankfurt cases discussed in the last chapter, I may intend to do X rather than Y, even though it is the case that unbeknown to me I would be blocked from doing Y if I opted for it instead..

(15.) In accordance with this account, my individual intentions are necessarily directed to what I can do; they have essentially centered contents, as it is sometimes put (Perry 1979; Lewis 1983b, Ch. 10). I can intend that I do such and such—or, equivalently, I can intend to do such and such—but I cannot intend, except in the special case of a supervisor, that such and such should be done. Joint as distinct from individual intentions will be discussed in the next chapter.

(16.) The word “desire” and the like—“prefer,” “want,” “seek,” and so on—are often used to express intentions. The claim, made later in the text, that I can pledge an intention but not pledge a desire introduces considerations that might suggest that there is a sharp distinction between desires and intentions.

(17.) I ignore here the limited expressive possibilities that might be illustrated in an utterance such as “Would that q!” or “Oh to make it the case that q!”.

(18.) In the ascriptive avowal of a belief or other attitude, it is worth noting that I do not just communicate that I have, say, the belief that p or the desire for R; I also communicate that I have the belief that I hold that belief or harbor that desire. But while I give expression to that higher-order belief, to return to an earlier observation, I will not strictly avow it. This is because it is not for the sake of communicating the belief more credibly, only because I have little or no option in the matter, that I choose to express the belief rather than report it.

(19.) Our observations on the avowal of non-credal attitudes bear indirectly on a familiar debate in metaethics as to what is the relationship between a moral attitude of approval or disapproval and an utterance that communicates the presence of that attitude: say, “I approve of X,” or “You ought to do X,” or “X is obligatory.” In their simplest forms, one of the standard approaches suggests that this sort of utterance expresses the attitude in the way that an assertion that p expresses a belief, another that it reports the attitude in the way in which an assertion that it seems to me that I believe that p might report a belief. The first of these alternatives is expressivism, which we discussed briefly in the first chapter, and the second is often described as subjectivism (Moore 1911). Neither is satisfactory, however (Jackson and Pettit 1998). In ignoring the role of the belief, simple expressivism would fail to explain why ethical utterances are voluntary acts of communication. In ignoring the difference between reporting and avowing, simple subjectivism would fail to explain why the utterance forecloses the misleading-mind excuse and helps put the speaker on the hook for any failure to live up to the attitude.

(20.) An assumption made throughout this study is that as data determinatively constrict belief, a point we mentioned earlier, so desiderata determinatively constrict desire and intention. Like the assumption about data, this is one of convenience to the extent that the main claims would survive, albeit in a

somewhat modified form, if it were dropped: if it were the case that the desiderata present in any instance may just constrain the alternatives available for me to desire or intend, not reduce them to one.

(21.) As in the case of belief, of course, there may be some cases where I would say that I changed my mind. But those would be cases where, as I think of it, the desiderata seemed to support the change—wrongly, as I may later judge.

(22.) This discussion ignores time-relativity as distinct from agent-relativity. But despite this restriction of focus, the examples of needs, tastes, and cravings indicate that desiderata may not just include properties that can make a prospect attractive, even if its realization will come at a time when the desire it currently elicits is no longer there—say, the property of enjoying posthumous fame; they may also include properties that make a prospect attractive only so long as the desire it currently elicits—the hankering it relieves—remains in place: say, the property of satisfying the yen for a cigarette. On these matters, see Parfit (1984); Pettit (2006b).

(23.) A similar example in the philosophical literature is Warren Quinn's (1993) "radio-man": a character who is disposed to turn on radios but without any even minimally intelligible rationale; certainly without any that he can invoke.

(24.) In helping a friend, of course, his or her identity will also matter to me. I will be fastened on them in particular, albeit under their aspect as a friend. On related questions, see Pettit (1997b, 2015c).

(25.) For a critique of particularism on these general lines—and for a critique of the closely related doctrine I call "interpretivism,"—see Pettit (2015c). For a deeper-running complaint about particularism, see Jackson, Pettit and Smith (1999). The line taken in the text would be consistent with thinking that properties can have an aggregate weight in combination that is not just a function of their individual, independent weights; in that sense, it can be moderately holistic. The approach is inconsistent, however, with the radical, particularistic holism that would not recognize any role for individual weights.

(26.) This will be so, strictly, only insofar as the intention does not cause the behavior by a deviant chain; it causes it robustly over cases in which the behavior is required by the intention. See Pettit (2018a).

(27.) I am grateful for a discussion of this issue with Pamela Hieronymi and Jay Wallace.

Access brought to you by: