Discovering Desirability

Philip Pettit

DOI:10.1093/oso/9780190904913.003.0007

Abstract and Keywords

In the ordinary world, we identify the desirable as something that is grounded in other properties, may diverge from what we desire, and, other things being equal, has a claim to govern what we desire. While desirability comes in many modes, moral desirability is grounded in relatively unrestricted considerations and enjoys a certain authority in resolving conflicts. Being creatures who avow and co-avow our desires, we are likely to find those desires diverging occasionally from our actual desires, and commanding our allegiance in the case of a conflict. Thus, we will begin to think of that which attracts avowal, being supported robustly by relevant desiderata, as having the governing role of the desirable. But as there are different modes of avowal, each supported by different sorts of desiderata, some neutral, some agent- or group-relative, there will be different and conflicting modes of desirability—this, by contrast with credibility. And the need to unify our own judgments of desirability into a single judgment of overall desirability, together with the need to universalize desirability so that it is standardized across individuals, will lead us to generate a notion of multi-lateral desirability that corresponds well with the ordinary notion of moral desirability.

Keywords: Credibility, desirability, credibility, multi-lateral desirability, agent-relative, group-relative

According to the argument in the preceding chapters, a simple, reportive community like Erewhon would not be a steady or stationary society. It would contain within itself the seeds of its own transformation, providing us as members with the means, the motive, and the confidence to take us beyond merely giving reports on ourselves and our attitudes. On pain of having few

excuses for failure, we would back ourselves to live up to certain self-ascribed attitudes and commit ourselves, in a strategic sense of the term, to those attitudes. Our commitments would include individual avowals and pledges of attitude, as well as co-avowals and co-pledges that we make in company with others.

Nothing in the developments reviewed so far takes us into the realm of ethics. At the point reached in the narrative, we do not yet make judgments of desirability, or indeed any prescriptive judgments, and we do not hold one another responsible for living up to them. The challenge now is to carry forward the narrative and show why the commitments that we make in avowals and pledges are liable and indeed likely to take us into prescriptive and ethical space. This chapter argues that we are in a position where it is natural to begin to think in terms of desirability, in particular moral desirability. And the next chapter shows that having come to make judgments in that mode, we are also going to be in a position to hold one another properly responsible to such judgments.

The argument in this chapter is that once we can avow and co-avow belief, we will be able to access the concept of credibility and, more pertinently to the purpose of the narrative, that once we are able to avow and co-avow desires, we will be able to access concepts of desirability, including the concept of moral desirability. And with that conceptual **(p.150)** access, we will be able to make suitable judgments of desirability and allow them to guide us in action.

The notion of a judgment that makes an appearance here can be equated with the act of making up your mind that was discussed in chapter 3. Whenever you avow a belief that p, claiming to be able to foreclose the misleading-mind excuse, you rely on having made up your mind that p. For present purposes, a judgment that p—and, to anticipate, a judgment that something is desirable—is just the act of making up your mind that p. While it is associated with the individual avowal of a belief, it may presumably materialize internally without an avowal actually ensuing. In this sense, you may make a judgment that p without asserting a word.

The argument of the chapter develops in five sections. The first analyzes the notion of moral desirability in order to make clear what the narrative is required to show. The second looks at how the social world is bound to appear to us from within the standpoint of avowing belief and desire, and the third at how it is likely to appear to us from within the standpoint of co-avowing those attitudes. The upshot of those discussions, as charted in the fourth section, is that while such perspectives support a single concept of the credible, they provide us with access to a variety of desirability concepts. The fifth section then argues that in face of that variety we are more or less bound to evolve a concept of the

desirable that integrates or transcends such different standpoints and that this concept is effectively equivalent to the concept of the morally desirable.

5.1 Desirability characterized
Desirability, moral and non-moral

That something is desirable may be taken to mean that you are permitted to desire it or, perhaps more strongly, that desiring it is recommended. But here it will be taken to mean, more strongly still, that you ought to desire it: that desiring that alternative is prescribed, **(p.151)** not just permitted or recommended. The fact that you judge that something is desirable in that prescriptive sense usually presupposes that it is one option in a set of alternatives, and in judging it to be desirable you rank it above the others; it is not just desirable in a generic way but, specifically, more desirable than those other options. The option that counts as most desirable may be a simple option like X or Y or Z in a three-way choice or, in the case of a tie, a disjunctive option like X-or-Y; in this case, some independent factor—perhaps just a mental toss of a coin—may be allowed to determine which disjunct to realize.[1]

How to distinguish moral desirability from other forms of desirability? Three features characterize it broadly, as noted briefly in chapter 1. The first of these features is that judgments of moral desirability are fixed by considerations that are relatively unrestricted in the range of interests served; they contrast on this front with judgments of prudence that target a particular person's interests, for example, or the judgments of patriotism that target those of a particular country. It is a mark of moral desirability, shared with certain other forms of desirability as well, that the features that determine whether an alternative is suitably desirable are not restricted to particular interests and beneficiaries.

The second mark of judgments of moral desirability is that they are grounded in features that are relatively unrestricted in standpoint, not just in the range of interests involved. Other concepts of desirability are generated by practical standpoints like those of law and etiquette, or by the epistemological standpoint from within which it is desirable **(p.152)** to believe that which the evidence supports, or by the standpoints associated with a set of projects—perhaps self-interested, perhaps disinterested—embraced by a certain individual or group. The concept of the morally desirable is not tied in the same way to such perspectives. This shows up in the fact that human beings routinely transcend such standpoints in moral judgment, debating the moral desirability of standing by a certain practice of law or etiquette, of allowing only the evidence to determine our attitudes, or of seeking to advance this or that project.

These comments bear on the first two marks of judgments of moral desirability: that they are grounded in features involving a relatively unrestricted range of interests and a relatively unrestricted standpoint. The third distinctive feature is that not only are judgments of moral desirability relatively unrestricted in their

grounds, they are also relatively authoritative. Judgments of moral desirability are typically assigned a weight that allows them to adjudicate between judgments of desirability dictated by rival interests and standpoints.

There are likely to be many judgments of desirability that conflict with one another because they reflect different ranges of interests or different standpoints. The third mark of moral desirability is that it is taken to be sufficiently authoritative to clinch the issue of what should be prescribed in such conflicts. Or at least to clinch the issue of what should be prescribed for ideal subjects. For, to anticipate discussion in the next chapter, what a judgment of moral desirability holds out as the preferred alternative may be beyond the capacities of relevant agents to the point where they are not fit to be held responsible for a failure to adopt it; the alternative, as it is often said, may be supererogatory, not a matter of obligation.

Is moral desirability equivalent, then, to overall desirability? No, because by most accounts the considerations relevant in moral judgments are ones that bear on certain conflicts; according to perhaps the most salient account, they bear on conflicts between the interests of different persons. That something is overall desirable, then, does not necessarily entail that it is morally desirable; considerations that give it this overall status may not be relevant to conflicts of that sort. And that something is morally desirable, being supported by considerations relevant in **(p.153)** conflicts between persons, does not necessarily entail that it is overall most desirable.[2]

By some accounts, the conflicts relevant to issues of moral desirability are restricted to conflicts affecting different human beings, by others they may extend to conflicts affecting any sentient beings, perhaps to conflicts where non-sentient nature or the divinity is involved, and even to conflicts, usually cast as prudential, between the interests of a person at different stages of their life. It is not necessary to decide between these accounts at this point, although it is worth noting that in the narrative developed later, conflicts affecting human beings or persons are given a certain priority.

The morally desirable and the morally obligatory

The concept of the morally desirable plays a central role in ethics or morality because of its connection with the more frequently invoked notion of moral obligation. But the connection may be understood in either of two ways.

The morally obligatory option, on one pattern of usage, is identical with the morally most desirable alternative, simple or disjunctive. But on the pattern to be adopted here, it is the morally most desirable option that it would be wrong or blameworthy for the agent not to take. On this construal, the morally most desirable option of all will not be the obligatory option if it counts as

supererogatory: that is, if it is so demanding that regardless of its desirability, it would not be appropriate to blame the agent for failing to take it.

Why not identify the obligatory option with the morally most desirable of all the options rather than with the most desirable option among **(p.154)** 'erogatory' alternatives—that is, among alternatives that would not be supererogatory choices? Either equation would work for purposes of the narrative, but it makes more sense to let obligation be understood on the second pattern. This construal has the advantage of marking clearly the accepted distinction between an option that is right in the strict sense of being morally obligatory and right in the wider sense that allows it to be supererogatory.

On this way of construing the notion of obligation, it is impossible to give an account of how we might get to make use of the concept in Erewhon, prior to having an explanation of how we might get to hold one another responsible for how we perform. The concept of obligation can only emerge properly, then, at the point in the narrative where it becomes intelligible why we should get to hold one another responsible, and the narrative will reach that point only at the end of the next chapter.

At that point, we will be able to understand, not just why certain options are obligatory but also why others are prohibited or permitted. An option will be prohibited if it is obligatory to avoid it, and an option will be permitted if it is not obligatory to avoid it and, at least in the sense in which "permitted" means "merely permitted", not obligatory to take it. Up until the point at which the concept of the obligatory is introduced, however, the focus will be exclusively on the concept of desirability, in particular moral desirability.

What role does judging that something is desirable, and in particular morally desirable, play in our thinking? There are three generic constraints that all judgments of desirability must satisfy and two specific constraints that judgments of moral desirability must satisfy in addition. The generic constraints reflect the role that any judgments of desirability must play in relation to desire, paralleling the role that judgments of credibility generally play in relation to belief. The specific constraints reflect assumptions about what judgments of moral desirability in particular should be taken to assert; whether the constraints are also satisfied by other judgments of desirability is not an issue that will be addressed here.

(p.155) Three generic constraints

The first generic constraint on judgments of desirability is that the desirability of any possible scenario relative to alternatives is grounded in the independent features of the alternatives on offer: ultimately in features that do not themselves have any prescriptive significance. As it was said earlier, the desirability of a scenario is fixed by how far it satisfies suitable considerations,

restricted or unrestricted. That scenario cannot cease to be more desirable than competitors without a change in the distribution of independent properties across alternatives; fix those properties, and the relative desirability of the alternatives will be fixed, too.

Why believe in the supervenience of desirability on other properties—its grounding in other properties—as this constraint is often described? The answer is, because it is encoded in the ordinary use of language. When I hold one alternative to be more desirable than another, it is always appropriate to ask about what makes it more desirable: what distinguishes it in independent terms from the other alternatives. And that question is appropriate only on the assumption that desirability is grounded in independent properties.

The second generic constraint on desirability judgments is that it is always possible that while one alternative in a choice is desirable—and even while I judge it to be desirable—I actually desire another. This may be because I am subject to any of those familiar influences that can disturb the effect of the desideratum registered in the judgment: say, the influence of impulse or whim. This constraint scarcely needs defending, since conflict of that sort between judgments of desirability and actual desires is a datum of common experience.

The third constraint is that in any such case of divergence, it is going to count as a rational failure on my part—a failure to function properly—if, other things being equal, I act on my desire and against my judgment of what is desirable. Other things will not be equal, if I do not have the capacity to do what I judge to be desirable, for example; or if there are other judgments of desirability still in play; or, to go to an **(p.156)** extreme possibility, if the judgment does not catch and I actually hold the contrary belief.[3] Absent such possibilities, however, the idea is that I will not function properly if I fail to let the judgment of desirability govern what I do. The idea is plausible, since it will be perfectly reasonable to ask me to explain myself in any situation in which I fail in that way.

These constraints may be named after what they impose or allow: grounding in the first case; divergence in the second; governance in the third. As they apply to any form of desirability, so they apply to moral desirability in particular and, by extension, to the obligation that it makes it possible to define.

The first, grounding constraint shows up in the fact that if I am told that one option is morally desirable, and another not, it always makes sense to ask about what is the difference—the independent difference—between them. The second, divergence constraint is reflected in our pervasive sense that we may often desire what is not morally desirable, even what we judge to be not morally desirable. And the third, governance constraint applies with particular force, since the judgment of moral desirability does not leave room for the idea that it is no more authoritative than certain rival judgments that remain in play. The

constraint means that when we judge something to be morally desirable, and things are otherwise equal, then it ought to have the role of guiding us, and if necessary correcting us, in the formation of desire and intention.

Two specific constraints

Where the three generic constraints reflect the role that judgments of desirability in general are expected to play in relation to desire, the two specific constraints reflect an assumption about what judgments of **(p.157)** moral desirability should be taken to assert. The assumption reflects widely supported intuitions about judgments of moral desirability, but it also has a methodological appeal, since it makes the exercise on hand more difficult rather than less difficult to complete. The assumption raises rather than lowers the bar to be crossed in providing a plausible explanation of how we residents of Erewhon could come to master and apply the concept of the morally desirable.

The first of the specific constraints is that when I judge that one among a set of alternatives is morally desirable—when I assent to the proposition ascribing such desirability to it—the property that I ascribe is not the property of being morally desirable$_{me}$, where this is distinct from the property, morally desirable$_{you}$, that you would ascribe if you were the one assenting to the proposition. The constraint is that 'morally desirable' is not indexical in the manner of 'mine' or indeed 'now'; it does not assume a different referent, depending on the identity of the utterer or of the context of utterance. Thus, when I say that it is morally desirable for a person to do something and you deny that that is morally desirable, we are not talking past one another, addressing different properties in our respective claims.

Where the first constraint holds that judgments of moral desirability do not vary in content as between speakers, the second holds that neither do they vary in truth-value by virtue of being about moral desirability. The first constraint is that you and I address the same proposition when, given the same context, I say that something is morally desirable and you deny this. The second is that in such a case, at most one of us is correct about that proposition. It cannot be that from my standpoint as an assessor—from the position that my evidence gives me—the alternative at issue truly is morally desirable, and from yours it truly is not; if it is morally desirable from one position, it is morally desirable from all. There may be nothing incoherent about the claim that truth-value may be assessor-sensitive, so that a given proposition should be deemed true from within one position of assessment and false from within another (MacFarlane 2014). But, so the second constraint holds, this is not the case with propositions about moral desirability as such.

(p.158) The two specific constraints satisfied by the notion of the morally desirable enable it to play a communal role in identifying a property such that people are capable of recognizing it in common, of taking their own guidance

from it, and of using it in critique of one another's performances or proposals. The concept is tailored by its satisfaction of the constraints to assume an important coordinating part in regulating people's relationships with one another.

The fact that the concept of the morally desirable satisfies the two specific constraints implies that the concept of obligation in the sense endorsed here—the concept of the most desirable alternative among 'erogatory' options—satisfies them, too. And that implied claim is independently plausible. If the obligatory is to serve its characteristic community-wide role in assessing options and actions, and in determining the responsibility of different agents, then it must be non-indexical and non-relative; it must allow different people to address the same content on the basis of the same criteria of assessment.

Given this understanding of what it is to judge that something is morally desirable, it is possible to explore how far we, the people of Erewhon, equipped with the committal practices of avowal and pledging, are likely to come to form such judgments. The argument to be offered is that making avowals and co-avowals—in particular, avowals and co-avowals of desire—is going to provide me and others in the community with a perspective from within which it is natural to begin to think in terms of the desirable and in particular of the morally desirable. Pledges do not figure much in this account, but they play a crucial role in the next chapter, helping to explain why it is also going to be natural for us to hold one another responsible to certain standards, including standards of moral desirability.

5.2 The view from within avowal
The robustly persuasive and attractive

When I speak for myself in Erewhon, making a potentially expensive avowal of a belief or a desire—in particular, the sort of belief or desire **(p.159)** that is likely to be relevant in continuing interaction with others—I rely on a solid basis for holding the belief or desire, and I guard against the effects of distractors and disturbers. The basis for belief is provided by the data I take care to survey, the basis for desire by the desiderata I carefully register. Given that I guard as needed against distraction and disturbance, data that argue that p elicit a relatively dependable disposition to act and adjust as if it were the case that p; and the desiderata that draw me to R elicit a relatively dependable disposition to act and adjust as if R were indeed attractive. I may later change my mind about whether it is the case that p or whether R is really attractive, as other data or desiderata come into play. But short of such an alteration, I will continue to believe that p and desire R. Assuming that the data or desiderata remain in place, and that I remain a competent subject, I will hold by that belief or desire robustly over variations in other factors, in particular variations in the distractions or disturbances that might affect me.

Given that basis for confidence about the belief or desire, I step out of the contingencies of the here and now when I avow the attitude. Taking the basis in data or desiderata as sufficient to elicit a dependable attitude, at least given a guard against distraction and disturbance, I treat the belief or desire as something I can stand by with relative assurance. I treat it as firmly enough entrenched for me to be able to put aside the possibility of appealing to a misleading-mind in order to excuse a miscommunication.

No matter how effective my protection against distraction and disturbance, of course, I still have to recognize that I may occasionally fail to display the attitude ascribed. Thus, persuaded by the data to avow that the gambler's fallacy is a fallacy, I still have to recognize that regardless of taking precautions against disturbers, I may lose sight of this truth in the excitement of the casino, and that if I do, I will not be able to excuse myself by saying that I changed my mind. Again, persuaded by the desiderata at hand to avow the desire to tell my friends the truth about some embarrassing episode, I may still have to recognize that, regardless of how far I guard against disturbance, the shame of telling the truth face to face may inhibit me from owning up to the episode with **(p.160)** someone who is particularly judgmental: and this, without being able to excuse my failure by a change of mind.

Think now about how I must view such disturbers, when in the wake of a failure I have to admit that they caused me not to live up to my avowed attitude. Given that disturbers, by assumption, are factors against which I can guard, I cannot claim not to have been able to act on the belief or desire or intention in play. But given that I did not change that attitude, I cannot claim that how I acted revealed my true mind. Hence only one alternative remains open. I must disown the action that I took, whether that involved placing a heavy bet on red after a run of blacks, or beating a hasty retreat from meeting with a judgmental friend. I must hold that that action does not reflect who I am. I must present it as the product of contingent influences that I do not identify with: influences that are not expressions of the mind for which I speak.

If I am disposed to take this view in retrospect, that also has implications for the view I must take in advance of any failure. It means that as I avow the attitude in question, backing myself to live up to it, I must not only hold the attitude avowed and be aware of holding it. I must also assume that I hold the attitude as a result of the impact of relevant data or desiderata, not as a result of any other influence, like that of a disturber or indeed a distractor. If I thought that my holding it was the effect of such an influence, then I would not have the confidence required for avowal.

It follows that when I hold an avowed belief that things are thus and so, for example—when I find that scenario relevantly persuasive—I do more than hold by the simple belief that they are thus and so. I hold also by the sophisticated

belief that the data support my believing that things are thus and so; it is not because of a lack of evidential care in guarding against distraction, or a lack of executive care in guarding against disturbance, so I assume, that I am led to believe that they are that way. In other words, I hold by the simple belief under the assumption—in general, no doubt, a default rather than a confirmed assumption—that there is nothing suspect at its origin: no distraction or disturbance. If I thought that there was, after all, then that would give me pause about **(p.161)** avowing it; I could no longer have the confidence to bet on myself to stick with it.

The same line of thought applies with the desires that I avow. When I hold an avowed desire that things be thus and so—when I find that scenario relevantly attractive—I do more than enjoy an attraction to their being that way. I enjoy that attraction but hold at the same time by the belief that relevant desiderata ground the attraction; that the attraction is not due to any lack of evidential care in guarding against distraction or any lack of executive care in guarding against disturbance. I stand by the attraction, perhaps letting it shape my actions, under the default or perhaps confirmed assumption that there is nothing suspect at its origin. If I thought that a lack of care played a part in the attraction, then, as in the case of belief, that would give me pause about avowing the desire; I could no longer have the confidence to bet on myself to stick with that desire.

Enter prescriptive concepts

These observations imply that, like everyone else in Erewhon, I have to treat the data on which I rely in avowing belief, and the desiderata on which I rely in avowing desire, as enjoying a certain privilege. Certainly, I have to do this in the case of the expensive avowals in focus here. Insofar as I am invested in the practice of avowal, and more generally in the role of spokesperson for myself, I have to see the data and desiderata as generators of attitude that I can safely mobilize. I have to see them, indeed, as the only generators of attitude on which I can rely in avowals, contrasting them with the contingent, erratic factors that may enter as a result of a lack of evidential or executive care. They represent the forces that must hold sway within my psychology—and the only forces that must hold sway there—if I am to be a reliable spokesperson for myself.

What am I to think, then, if I realize that the relevant data support a belief that p but, for whatever reason, I do not hold this belief, or even hold by the belief that not-p? What am I to think if I realize that the relevant desiderata support the desire for R but, for whatever reason, I do not harbor this desire, or even harbor a desire for not-R?

 (p.162) As someone committed to speaking for my attitudes, and to holding those attitudes for which I speak, I have no option but to think in each case that by a principle implicit in that practice, the actual state of my belief or desire is not what it should be. And equally, I will have no option but to seek, if possible,

to let the relevant data or desiderata elicit the missing attitude instead. It may not be futile to seek this result. The recognition of my failure—the recognition that I cannot claim to speak for myself, if the actual state of my belief or desire remains as it is—should supplement the relevant data or desiderata in eliciting a change of attitude. It should activate a natural desire to give careful attention to those factors, guarding against distraction and disturbance.

But not only must I take a critical view of a current failure that I recognize in the formation of my attitudes; I must also take a critical view of a past or prospective failure on my part or indeed of any similar failure on the part of another person. Thus, suppose I now recognize that I formed a belief or desire in the past that, because of a lack of evidential or executive care, was not supported in the manner appropriate to the practice of avowal—in the manner that made it fit for avowal. In that case, I must judge that my attitude was not then as it should be. And suppose I recognize that despite being committed to avowing it, you hold a belief or desire that is not supported by data or desiderata in an avowal-fit manner. In that case, too, I must think that your attitude is not as it should be.

The upshot is that as commissive creatures who avow our beliefs and desires to one another, claiming to be able to speak for such attitudes, we must treat our beliefs as attitudes for which we can speak when, and indeed only when, the propositions believed are persuasive in a way that makes them fit for avowal; and we must treat our desires in the same way when and only when the prospects desired are attractive in a parallel avowal-fit manner. This is going to be a matter of common awareness in Erewhon, given that the importance and price of claiming to be able to speak for ourselves is obvious: the evidence supporting the claims is salient for all, the evidence that that evidence is salient is itself salient for all, and so on in the usual pattern (Lewis 1969).

 (p.163) The content of these claims will find natural expression among us in talk of what I or anyone else ought to believe or desire. What I ought to believe qua someone who avows beliefs—qua someone who claims to be able to speak in avowal for myself—are just those propositions for which there are data enough to elicit belief, when I exercise any evidential and executive care that may be needed. What I ought to desire qua someone who avows desires are just those prospects for which there are desiderata enough to elicit desire when I practice similar forms of care. And I must acknowledge those claims about what I ought to believe and desire as general truths that hold in abstraction from the precise identity of the relevant data or desiderata.

Thus, the persuasive, as determined within the practice of avowal, will present as what I ought to believe; the attractive, as determined within the practice, will present as what I ought to desire. That which is persuasive and attractive within this practice is going to be persuasive and attractive robustly over variations in

any factors other than the relevant data or desiderata. In particular, it is going to be persuasive or attractive over variations in the distraction and disturbance that evidential and executive care are designed to protect me against. It will be robustly persuasive, as it may be put, or robustly attractive.

It will be plausible for all of us in Erewhon, then, that we each ought to believe that which is robustly persuasive, desire that which is robustly attractive, and avoid beliefs and desires that fail these conditions. And this will be plausible as an abstract truth of principle, so that we must admit that we ought to respond in this way to anything, and only to anything, that proves to be robustly persuasive or attractive for us, even if we do not currently recognize that it is persuasive or attractive.

This holds of us, as mentioned, insofar as we are creatures who are invested in speaking for ourselves in avowal of our attitudes. Is that a significant restriction on the principle? Not really. At least not, on the assumption that as we learn to treat others as interlocutors, and find that they treat us in the same way, we will treat ourselves as potential interlocutors to ourselves. According to this assumption, having learned what it is to be on the hook with others, we invent a parallel **(p.164)** hook on which we will find ourselves if we self-ascribe a desire or belief for which we can see no robustly supportive data or desiderata.

The assumption means that even beliefs and desires that are irrelevant in our relationships with others, even indeed beliefs and desires whose absence in us others are not in a position to test for, will fall under the principle that lets the robustly persuasive guide belief and the robustly attractive guide desire. Thus, to revert to earlier examples, the principle will govern the avowal of a belief that I once saw a mathematical equation in the clouds, a belief that Mogli is probably honest, a desire to fly, or an intention, if p or q or s, to X.

Is the assumption plausible? Yes, because the categories of the robustly persuasive and attractive are applicable to all beliefs and desires, and I am in a position to apply them in my own case, even when others cannot readily do so. But why apply them in my own case? Because doing anything else would involve treating like cases differently. It would involve a failure in my own case to invoke categories that I invoke in relation to others, expect others to invoke in relation to me, and treat as relevant when others invoke them in that way. It would mean assuming the role of interlocutor in relation to others while rejecting that role in relation to myself.

This assumption will not play a crucial role in the evolving narrative, but it may be adopted as a plausible generalization; it will be relevant to the discussion of the linkage between morality and personhood in the first section of the final chapter. Not only should the beliefs and desires that are relevant in our relationships with others be guided by the robustly persuasive and the robustly

attractive, then; all our beliefs and desires, even ones that do not figure in our avowals to others, should be guided in this manner.[4]

With the developments charted in this discussion, we in Erewhon have entered prescriptive space. We are in a position to access the idea of what I or anyone else ought to believe—what is credible for me—and **(p.165)** the idea of what I or anyone else ought to desire: what is desirable for me. And those concepts are prescriptive insofar as they satisfy the grounding, divergence, and guidance constraints mentioned earlier.

First, whether something is credible or desirable relative to me is going to be grounded in its relations to data or desiderata; these will explain why a proposition is credible, a prospect desirable. Second, what is credible and what I judge to be credible may diverge from my actual beliefs, what is desirable and what I judge to be desirable may diverge from my actual desires, since distraction or disturbance may play a role in generating my actual attitudes. And third, assuming other things are equal, it would be a functional failure not to let what I judge to be credible govern or determine what I believe, or not to let what I judge to be desirable govern what I desire. It would involve an inconsistency with what I assume in following the practice of avowal.

Other things will not be equal, and this third lesson will not follow, as mentioned earlier, if the judgment of credibility or desirability is not the only one in play. This observation is important, as it turns out that there are judgments of credibility and desirability that answer to practices of co-avowal as well as the judgments of credibility and desirability that answer to the practice of individual avowal. In order to mark this contrast, what is credible or desirable in light of the practice of individual avowal will be cast henceforth as what is individually credible or desirable, leaving open the possibility that it is not credible or desirable in other ways.

Avowals of individual credibility and desirability

With access to the concepts of the individually credible and individually desirable, I and others in Erewhon can form beliefs to the effect that something is credible or desirable in that way. And, of course, we will avow such a belief in asserting that something is individually credible or desirable; that assertion will communicate the belief pragmatically rather than report on its presence. Avowal is a potentially recursive operation such that we may avow a belief in a content— that something is credible or desirable in some way—whose very availability to us as a **(p.166)** content to be believed itself presupposes the prior use of avowal. While the standpoint of avowal enables us to gain access to the concepts of the individually credible and desirable, applying these to what we find robustly persuasive and attractive, it enables us at the same time to avow beliefs in propositions that ascribe those very properties of credibility and desirability.

How in Erewhon might I avow a belief in the individual credibility of a proposition 'p' or in the individual desirability of a prospect R? The usual linguistic devices will be at my disposal. I may express such a belief, as just noticed, by asserting that it is credible that p, or that R is desirable. But I may also self-ascribe such a belief, and still retain the force of an avowal, by saying that I believe that it is credible that p or that R is desirable. And equally I may resort to remarks that serve in context to explain, not why I believe that p or desire R—I may not actually do so—but why it is credible that p or why R is desirable: I may say, for example, "The data stack up in support of 'p,' " or "R would be a lot of fun."

These observations show that, like others in Erewhon, I would naturally be led, just in virtue of making personal avowals, to develop a prescriptive viewpoint on myself. I cannot practice avowal without privileging a robust personal standpoint, as it may be described: the standpoint in which I am responsive to the robustly persuasive in the case of belief, and to the robustly attractive in the case of desire. This standpoint is ideal in the sense that, by assumption, it is free from those distracting or disturbing factors that might warp the influence of suitable data on belief, suitable desiderata on desire. And so, assuming that standpoint, I can prescribe for how my actual self ought to perform.

I can prescribe that actually I ought to stick with a belief that the gambler's fallacy is a fallacy when I go to the casino, or that actually I ought to speak truthfully in face-to-face meetings with my friends. And, should it prove impossible to guard against such disturbing influences, I can prescribe that I ought to avoid occasions when they arise. I ought to avoid exposing myself to the disturbers, as I ought to avoid exposing myself to the distractors, that undercut the effect of data or **(p.167)** desiderata. I ought to avoid visiting the casino, or I ought to avoid difficult face-to-face encounters.[5]

5.3 The view from within co-avowal

My individual perspective in avowal lets me identify the robustly persuasive and attractive, and leads me to give it prescriptive status, treating it as representative of the individually credible, on the one side, the individually desirable on the other. But my perspective in co-avowal and co-acceptance allows me to do something parallel at the social level and complicates the concepts to which I and others in Erewhon will enjoy access.

The complication is exacerbated by the fact that co-avowal may be bounded or unbounded, as appeared in the last chapter, and that bounded co-avowal may take as many different forms as there are different bounds. In the discussion that follows, the unbounded case will be given priority, since conversation of this kind is tied up with individual avowal and, as noted, enjoys a certain inescapability for human beings. That discussion will address the co-avowal of belief first, the co-avowal of desire second, and in each the lessons it teaches will

be drawn first for unbounded conversation and co-avowal and then for bounded counterparts.

From within the unbounded co-avowal of belief

Suppose that I essay the co-avowal of a belief that p, opening up a potential, unbounded conversation with others. Suppose that some others go along, acquiesce in the co-avowal, and essay further co-avowals themselves. And imagine then that in an exercise involving various **(p.168)** episodes of rejection, rejoinder, and revision, we come into convergence with one another. The result should be that we will each endorse a set of beliefs that any one of us is in a position—indeed is manifestly in a position—to avow on behalf of all of us and, by aspiration, on behalf of any others who join up. In the domain explored, this exercise will reveal certain propositions as robustly persuasive for us all; and this, as a matter of common awareness from within the standpoint that we share.

It will be manifest to each of us in such a case that due to distraction or disturbance, anyone may occasionally fail to believe what is robustly persuasive within this group. But, recognizing what the interpersonally tested data elicit, we must each be disposed to disown any such failure: that is, to treat the distraction or disturbance as warping the performance required of us within the standpoint presumptively shared with an open number of others.

This means that what is robustly persuasive from the common standpoint of this open group is a prescriptive category on a par with what is robustly persuasive from an individual standpoint. The practice of co-avowal will require each of us to believe that which we take to be robustly persuasive for all and ready for co-avowal by any one of us. What is robustly persuasive in this way constitutes the commonly credible, as we in Erewhon might come to articulate it.

The prescriptive status of the commonly credible shows up in its satisfying the grounding, divergence and governance constraints listed earlier. What is commonly credible is grounded in the evidence that is common to an open group of others. What is commonly credible, and what I judge to be commonly credible, may diverge from what I actually believe under the influence of distracting or disturbing influences. And if my judgment of common credibility is sound, it would be a functional failure on my part not to let that judgment govern my beliefs. Or at least it would be a failure on the assumption that other things are equal, and in particular that there is not some rival judgment of credibility in play.

This last observation is important, since it appears that there is always going to be a rival judgment of credibility in play: viz., a judgment **(p.169)** as to what is individually credible rather than commonly credible. So how then are these judgments likely to compare with one another?

It turns out that they relate in a wholly convergent manner. What is individually credible for each of a number of people is bound to be commonly credible. The reason is that the data relevant in my perspective ought to be relevant in everyone else's as well, since evidence for one person is evidence for all. Let two people confront the same body of evidence, where this includes what they know by way of background as well as the data on hand, and it will make little sense to imagine that it might argue for a different belief on the part of each.[6] The data that make something individually credible for me, then, are bound to be able to serve as data that make it individually credible for you, and vice versa. And whatever is individually credible for each member in any such twosome, or in any open group whatsoever, is commonly credible for all.

As the individually credible is bound to be commonly credible, so for similar reasons the commonly credible is bound to be individually credible. All that it means, in effect, for something to be commonly credible is for it to be credible for any individual, no matter who. And if something is credible for any individual, it is going to be credible for me in particular and for you or anyone else in particular.

These observations show that since individual and common credibility converge, there is only one category in play. If something is credible for me or anyone else in particular, it is natural to say that that is because it is credible for no matter who; the general credibility explains the particular credibility, and not the other way around. Thus, the conclusion that individual and common credibility do not come apart may be cast as the claim that common credibility is a master category. This, as will appear later, marks a sharp contrast with the relationship between the individually and commonly desirable.[7]

 (p.170) As avowal is recursive, so, too, is co-avowal. Once the category of the commonly credible becomes available in Erewhon, we members can avow beliefs in propositions to the effect that it is commonly credible that such and such simply by asserting that such and such is commonly credible. Indeed, we will presumably be ready to co-avow that belief, since we will surely expect others to go along, provided they exercise any required evidential and executive care. Thus, we can co-avow a belief that something has the property of being commonly credible, despite the fact that property will have become accessible to us only in virtue of having practiced co-avowal with simpler beliefs: beliefs not involving the property in their content. And for reasons familiar from other cases, the fact that we can co-avow a belief in this way means that in suitable contexts we will also be able to avow it by ascribing it to ourselves or by explaining why it is true.

The standpoint from within which I believe—and avow the belief—that something is individually credible is idealized, as appeared earlier; it represents a standpoint from within which I can prescribe for my actual self. We now see

that the standpoint from within which I believe—and no doubt avow and co-avow the belief—that something is commonly credible converges with that standpoint. It represents the ultimate point of idealization from which I can prescribe matters of belief for my actual self. There is no tussle between the individually and the commonly credible, and no problem about their offering different guidance on what I should hold.

From within the bounded co-avowal of belief

As the commonly credible will become defined for the members of an unbounded group, so a counterpart ideal—the jointly credible, as it may be put—is likely to be defined for the members of any bounded **(p.171)** group—say, a group devoted to some cause or some creed, whether or not organized as a group agent. The considerations that made this plausible in the unbounded case will also make it plausible in this, explaining how we can avow and presume to co-avow a belief in the joint credibility of a proposition, whether by expressing it in an assertion of joint credibility, ascribing the belief to those of us in the relevant group, or explaining why it is true by our lights.

How does the commonly credible relate to what is jointly credible, now from within this group, now from within that? If we are given a motive within a bounded grouping to set aside some beliefs that we take to be commonly credible, that must be because of constraints that are independent of data. This may be the desire to find a compromise among a fixed set of members, including some who by our lights are not suitably attentive to the data. Or it may be the desire to stick with a certain core of doctrine, regardless of how far it outruns the presumptive data, even perhaps conflicts with the data.

If a consideration unrelated to data constrains the category of the jointly credible, however, keeping it apart from the commonly and individually credible, then that undermines any hold it can have on us. It means that insofar as it does not coincide with the commonly credible, the idea of what is jointly credible for us in a particular group directs us only to things such that we have to act as if we believed them, assuming we wish to maintain our connection with other members of the group. Thus, we cannot take the category seriously in the determination of belief.

There is a noteworthy difference in this respect between the bounded group that organizes itself for action, incorporating in the manner of a group agent—say, a voluntary association, a company, or a church—and the bounded group that does not. In each case, members may have to set aside some of the beliefs they themselves take to be commonly credible. But where only contingent pressures will have this effect in the unincorporated group, the requirements of group agency make it more or less inevitable in the incorporated case. Any group agent has to establish a coherent set of judgments designed to guide all its

members when **(p.172)** they act in the name of the group; and it will have to do this, inevitably, regardless of their particular opinions (List and Pettit 2011).[8]

When we speak here of the jointly credible—a similar lesson applies to the jointly desirable—what we have in mind is what is credible for each one of us, as members of a bounded group. And whether the group be incorporated or unincorporated, it should be clear that the notion of the jointly credible can have no serious hold on us as individuals. The idea of the commonly credible—the idea of what is credible in light of data available to all—retains its status as the master category. This, as will appear, marks a deep contrast between the idea of the credible and the idea of the desirable.[9]

From within the unbounded co-avowal of desire

We now turn from the co-avowal of belief to the co-avowal of desire. Suppose that I presume to co-avow a desire for R, aspiring to speak to an open audience in an unbounded conversation on the topic. And suppose that those who pay attention, at whatever time and place, acquiesce in that avowal, essay further co-avowals themselves, and come to identify a shared set of desires in an exercise involving rejection, rejoinder, and revision. Each of us at that point will be in a position—indeed manifestly in a position—to avow those desires on behalf of the group and, by aspiration, on behalf of others whom we allow to join us. Within the domain explored, this exercise will reveal certain scenarios as robustly **(p.173)** attractive for all of us: they will appeal to us in light of desiderata that we are each disposed to acknowledge, absent distraction and disturbance, from within the common standpoint we assume.

What sorts of scenarios are likely to prove robustly attractive, and fit candidates for co-avowal, from within this standpoint? The issue is more complex than with the robustly persuasive. What count as data for one count as data for all. But what attracts one person may fail, indeed fail with a certain inevitability, to be attractive from a standpoint shared equally with others. I may desire my daughter's welfare on the basis, precisely, that she is my daughter, where others will only desire her welfare as they might that of a random person. With such an agent-relative desideratum in play, what is robustly attractive from within my individual standpoint may not be robustly attractive from within a standpoint that I purport to share with others. From within my individual standpoint, my daughter's welfare may be robustly attractive, even at a cost to the welfare of other children; from within a common standpoint, the welfare of all children will presumably count in the same way.

Returning to the question raised, then, what scenarios are likely to show up as robustly attractive, and fit candidates for co-avowal, from within the standpoint of an open group? The answer is, those scenarios that are attractive in virtue of promising to realize agent-neutral desiderata: that is, desiderata we are each liable to care about in a similar manner. Plausible neutral attractors may make it

robustly attractive for all of us that social norms like truth-telling or non-violence should obtain; that our species should survive into an indefinite future; that there should be no unnecessary suffering; and so on.

We each have to recognize that we may fail to live up to what is robustly attractive from a common standpoint, due either to distraction or disturbance. Disturbers will include the self-centered preferences that may detach us from the common point of view as well as the wayward impulses that may affect any one of us individually. But when we take something to be commonly attractive, we must assume that, as we guard against distraction, so we are each going to guard against disturbance, disowning any desires that they might introduce and seeking to stay faithful to the shared standpoint.

 (p.174) This means that like that which is robustly persuasive for the open group, that which is robustly attractive for the open group is a prescriptive category, directing us to what is desirable from within a standpoint that we share with an indefinite number of others; this is the category of the commonly desirable, as we in Erewhon may think of it. The commonly desirable satisfies the grounding, divergence, and governance constraints associated with all prescriptive categories. It is going to be grounded in the desiderata—the agent-neutral desiderata—that make something robustly attractive to me and others qua members of an open group. It may come apart from what I actually desire—say, as a result of distraction or disturbance—even when I purport to occupy a common standpoint with others and to think as the member of an open-ended group. And assuming other things are equal, my identification with that standpoint rationally or functionally requires me to let my judgment of the commonly desirable guide or dictate what I actually desire.

But are other things likely to be equal? In particular, is the judgment of common desirability likely to converge with that of individual desirability? Not as a general matter. In any case, where I or you or anyone else makes a judgment as to what it is commonly desirable to do, we are each liable to make a different judgment about what is individually desirable. How, then, are these judgments likely to relate to one another? It turns out that they are liable to diverge and that neither has a clear priority in relation to the other.

The individually desirable and the commonly desirable may often coincide, of course. It may be both individually and commonly desirable, for example, that I should tell the truth to others; this may be robustly supported by desiderata relevant in each standpoint. But in many cases, these standpoints are likely to offer inconsistent prescriptions. The role of agent-relative desiderata in determining what is individually desirable means that what I would prescribe from an ideal, individual standpoint may diverge from what I would prescribe from the ideal, common standpoint. And of course, it also means that what I

prescribe from my ideal, individual standpoint may differ from what you prescribe from yours.

The consequence of this is that neither the individually desirable nor the commonly desirable can play the role of the unconditionally **(p.175)** desirable, at least not in any context in which both are relevant, as they will often be. The categories target what is robustly productive of desire, in the one case, under the characterization of robustness that goes with my practice of individual avowal, and in the other, under the characterization that goes with our practice of common co-avowal. Those perspectives may come apart in a way in which the corresponding perspectives in the case of belief do not. And so, neither can direct us toward a master category akin to the category of the commonly credible.

As avowal is recursive, so, too, is co-avowal. Once the category of the commonly desirable becomes available in Erewhon, then, we members are likely to form beliefs in propositions to the effect that this or that scenario is commonly desirable. And when we assert that a scenario is commonly desirable, we will avow our belief in its desirability rather than merely reporting on it. Others may be expected to go along with such an assertion, of course, if the scenario is commonly desirable and they are not evidentially or executively careless. And so, our assertion will naturally purport to have the status of a co-avowal of the belief, not simply an individual avowal.

As I may avow a belief in the individual desirability of a scenario by expressive, ascriptive, and explanatory devices, so I may resort to such devices in avowing or presuming to co-avow a belief in its common desirability. Depending on context, I can avow or presume to co-avow a belief in the common desirability of a prospect, R, by saying that it is desirable or commonly desirable; by saying that we desire it or believe that it is desirable or commonly desirable; or by explaining its desirability appropriately: say, by reference to how much fun it would be for everyone or to how it would give us each a fair return.

From within the bounded co-avowal of desire

According to the emerging narrative, I am naturally led by the practice of making individual avowals of desire, to develop one prescriptive point of view: a personal standpoint from which I can judge my actual performance, letting what I desire be assessed in terms of whether it is **(p.176)** individually desirable. And in the same way, I am led by the co-avowal and co-acceptance of desire that I practice in unbounded conversation, to develop a second prescriptive standpoint on desire: a common standpoint from which I can judge what I desire, letting it be assessed in terms of whether it is commonly desirable.

But this, of course, is not all. The discussion so far has focused on what is likely to count as attractive from the point of view of an unbounded group and on what

is commonly desirable in the sense of being robustly attractive from within the perspective of the group. But the argument developed in the case of that group supports parallel conclusions for this or that bounded group. As the perspective of the unbounded group will direct us to the category of the commonly desirable, so the perspective of any bounded group will point us toward the category of what is jointly desirable for members of that group in our part as members; this will be identified by what is robustly attractive to us in that role.

We the members of Erewhon, like the members of every society, are likely to find ourselves in any of a number of bounded groups; indeed, our own community, should we come to discover neighboring societies, would also constitute one example. And within such a partial grouping, as within the unbounded community imagined, we will each conduct conversational exchanges with others in which we co-avow and co-accept a range of desires that reflects the properties that matter from our shared standpoint, identifying scenarios that we will see as jointly attractive.

These properties will include group-relative properties that matter to us greatly as members—the welfare of our caste, the prosperity of our clan—but may not matter much to us in other roles. And so, for each such grouping, there is likely to be a notion of the jointly desirable that operates prescriptively but is in potential conflict with rival forms of desirability. It will satisfy constraints like those of grounding, divergence, and governance. And it will be a property such that any one of us can avow a belief in its presence, and presume to co-avow the belief within suitable contexts. We may do this by expressing the belief in an assertion that something is jointly desirable. Or we may avow it, for the same reasons as in other cases, by ascribing it to ourselves or by explaining why it is true.

 (p.177) The category of the jointly credible, as noticed earlier, is not of much significance, since it is only in virtue of constraints unrelated to data that the jointly credible might not coincide with the commonly and individually credible. The jointly desirable has a very different status, since it is group-relative desiderata, and not independent factors, that may force it apart from the commonly and the individually desirable. Thus, the divergence from individual and common forms of desirability, like the divergence between the individually and commonly desirable—and like the divergence between what is jointly desirable from the viewpoint of different groups—is not just salient but highly significant. This divergence in desirability, contrasting as it does with the convergence in credibility, is at the focus of the discussion in the remainder of the chapter.

5.4 A breakthrough and a shortfall

A conceptual breakthrough

If the argument so far is sound, then, in the wake of the developments charted, I and you and others in Erewhon will enjoy a conceptual breakthrough but suffer at the same time a conceptual shortfall. The conceptual breakthrough occurs in the areas of both belief and desire, the shortfall is confined to the area of desire alone.

The breakthrough is that we will become able to think in prescriptive terms, enjoying a position from within which we can distinguish between things as we actually believe or desire them to be and things as we ought to believe or desire them to be. How we ought to believe and desire things to be, in this way of conceiving of them, is how we would hold or want them to be, if we conformed to the constraints associated with a position we privilege: if we let the associated data or desiderata robustly determine the attitude. Depending on context, this is the position of the avowed self, or the self that is projected in one or another form of co-avowal. It is the position of the self as spokesperson for itself, now in one context, now in another.

 (p.178) It is a real gain for us in Erewhon to be enabled on this basis to think and talk in prescriptive terms, avowing beliefs as to what is credible and we ought to believe, what is desirable and we ought to desire. Once equipped with this capacity, there are two selves that we each confront: first, the self we project in our role as spokesperson for ourselves, more or less regularly living up to it; and second, the self we display in our actual behavior, when we fail to live up to the self for which we speak. As agents who speak for ourselves, we naturally identify with the bespoken self and, taking it as our point of view, look on the attitudes of the failing self as attitudes to reject. From the perspective of the bespoken self, we must each think that what that failing self thinks and feels and does is not me; it is not who I am.

In the life that we enjoyed prior to making commitments, we might have responded to incentives to prove reliable to others; we might have generated aggregate social patterns like those of general truth-telling; and we might have been in a position to recognize that result and to treat the patterns as social norms to guide ourselves by. And equally, in that life, we might have been in a position to be frustrated at our having failed to prove reliable, say, because of not guarding against distraction and disturbance, and at consequently losing out in the reputation stakes. But still, we would have felt that frustration from the point of view of a single agent or self. It would have distanced us from what we did but not from the beliefs and desires that those doings fail to reflect, and not in that sense from the self that we ideally are.

This would have been so even if it is assumed, plausibly (Tomasello 2016), that natural selection would have favored the appearance of ever more altruistic desires and ever more allo-centric beliefs. On this assumption, we would not

have needed to think strategically in order to establish mutual accommodation and reliance; we would have become spontaneously more disposed to take one another's welfare into account, and to adjust to one another's testimony. But no accretion of such desires or beliefs, however other-regarding it made us, would have provided for a critical, prescriptive viewpoint on ourselves. None would have given us an alternative self with which to identify (Pettit 2018b).

All of that changes once we begin to practice acts of avowal and pledging, and gain access to the concepts of credibility and desirability **(p.179)** that they bring on stream. Those shifts enable us to recognize that how we are may or may not be how we are committed under relevant practices to being, and that when we do not conform to the requirements of those practices then we display a sort of failure. We fall short in ourselves of the self we spoke for; we believe what is not credible, or desire what is not desirable, by the lights of that bespoken self. And when we recognize the actuality of failure, we simultaneously grasp the possibility and attainability of success. We see it as within our grasp: what we can become, if only we allow the bespoken self to help shape the self we actually are.

The perspective of the bespoken self is also, it should be noted, the perspective of the beholden self. For the self that we speak for in avowing or pledging, co-avowing or co-pledging, is a self that we have given others the right, under the rules of relevant practices, to expect us to display. If we do not display that self in action as well as word, then the rules of avowal or pledging give them the right to ignore certain excuses—to treat us as uncooperative parties—and consciously or unconsciously to impose associated reputational costs.

This conceptual breakthrough ought to be welcome in itself, opening up a wholly new way of thinking, and holding out the possibility of a new sort of personal aspiration and criticism. But it ought also to be welcome insofar as it is bound to serve our interest in being able to rely on others and to get others to rely on us. For with the extra resources available in any given context, we will each have an enhanced capacity to assure others of our reliability. I will be able not just to avow or presume to co-avow a belief that p or a desire for R, but to avow or presume to co-avow a belief that p is credible or that R is desirable. And in reaching for such an extra means of communicating my belief or desire, inviting you to rely on me, I will signal that I must pay an even heavier reputational cost, should I fail in the absence of excuse to live up to what I say.

A conceptual shortfall

But while the breakthrough into prescriptive space is a huge benefit for us, the people of Erewhon, it conspicuously gives us less in the area of desire than it does in that of belief.

(p.180) The fact that the commonly credible is a master category in relation to the individually and the jointly credible has two welcome results. First, I need not be divided within myself as a result of being disposed to avow or co-avow different beliefs when different perspectives are available; the only belief to which I am going to be committed as a general matter is the belief that I take to be commonly (and so individually) credible. And second, I need not be divided from others, at least not as a matter of necessity, as a result of being disposed to avow that belief; the belief to which I am committed on any relevant topic is a belief, being commonly credible, to which anyone, plausibly, should be committed. The first lesson is that I can be unified within myself in the domain of avowed belief, the second that I can take my avowed belief to be universally compelling, not just compelling in my particular case.

These implications of unity and universality are available in the case of belief because the data that are given a special role in avowal and co-avowal are the same data across those practices. They are the same data across the practices of individual, joint, and common avowal. And they are the same data across the different positions that you and I and other people may occupy. Not only do they enable me to form a single mind on the issues answerable to the data, they also enable me to make persuasive sense of that mind—specifically, of the beliefs I hold—in terms accessible to others.

Such unity and universality of perspective fail in the area of desire. This, as emphasized at various points, is because the desiderata that engage me in individual, joint, and common practices may vary considerably, as may the desiderata that engage you and me and others in our individual or joint practices. And that is so, despite the fact that we need not be exclusively self-centered in the desiderata that engage us in practices of individual avowal, nor particularly group-centered in those that engage us in practices of joint avowal.

The desiderata that move me in the individual avowal of desire, however altruistic I am, will certainly include agent-relative attractors—that the action will help my child or, more altruistically, enable me to keep a pledge—and may be the only set to do so. The desiderata that **(p.181)** engage me or you in joint avowal will certainly include group-relative attractors—that this will promote a shared interest—and may be the only set to do so; again, this will be so, however oriented we are toward the interests of others. And the desiderata that I mobilize in common avowal will include neither. They will be restricted to agent-neutral, group-neutral properties: properties like that of promoting truth-telling, reducing pain or ensuring peace.

As I may be moved by different desiderata in practices of individual or joint avowal, so you and I and other persons may be moved by different desiderata. Being differently motivated as individuals, we may weigh various desiderata differently, whether they be agent-relative or agent-neutral in character. And

being differently motivated as members of distinct groups, we may also give different weights to various group-relative or group-neutral desiderata.

The lack of unity in the case of desirability will show up wherever different perspectives are relevant on a decision I face, and different sets of desiderata compete for determining the desirable option. One of my alternatives in such a context may be individually desirable, another commonly desirable, and yet another desirable from the joint standpoint of some contingent grouping. In a time of need, for example, it may be individually desirable that I devote my efforts to my children, jointly desirable from the standpoint of my neighborhood that I devote them to the welfare of those who live nearby, and commonly desirable that I put them at the service of people in general.

In any such situation, it may be clear to me how I should choose under one hat, and clear to me how I should choose under another. And so, it may be clear to others that I will be ready to avow a desire for one option under the first hat, a desire for another under the second. But this will be of little use to me or them. I will want to determine which option is desirable for me, independently of the hat I wear. And others will want to know which option is desirable for me, independently of that hat. Otherwise, I will be inscrutable to both myself and others. I won't know, and they won't know, where I stand.

Even if we could resolve the problem raised by this lack of unity, the lack of universality in the case of desirability would still constitute a **(p.182)** difficulty. Resolving the problem of unity, it might be the case for each of us that in any conflict between what we find individually, jointly, and commonly desirable, there is always or generally a fact of the matter as to what is overall most desirable for us personally. But that would still leave the problem of universality in place; indeed, as will appear, it may make that problem even more challenging that it might have otherwise been.

The universality problem is that there is no shared currency of what is desirable for anyone—no accepted idea of multilateral desirability—that might enable us to present ourselves as disposed to do what is desirable in that sense. There is no currency that can serve on the side of desire in the way that the currency of common credibility can serve on the side of belief. Thus, we lack an idea of desirability that would enable us to support our commitments to behave in a congenial manner that is acceptable on all sides and to persuade others of our reliability on that front.

This universality problem will be made worse, if we have evolved a notion of what is personally most desirable, solving the unity problem. For this idea will naturally suggest the thought that we will each do what is personally most desirable. And that will be troubling in two ways. First, it may make it seem unlikely that we will reliably stick to the commitments we have made and behave

in a manner congenial to others. Second, even if this is not so, others won't be in a position to assure themselves readily that it is not so; it will be hard for them to tell what is personally most desirable for us in this or that situation.

An idea of multilateral desirability would enable us to get over both of these difficulties. By seeing that a certain option was multilaterally desirable— desirable on many, indeed all sides—others could be fairly confident that we will take that option: after all, it would be bound to appeal on our side as well as on theirs. And they would be able to see whether something is multilaterally desirable without conducting a psychological investigation of our particular dispositions; a given alternative would presumably be desirable for us in that way insofar as it would be desirable—presumably, more or less clearly desirable —for just about anyone in our situation.

(p.183) 5.5 Toward the morally desirable
How we could repair the shortfall

Is there a likely means of resolving these problems of disunity and non-universality? In particular, is there a way in which we might construct a concept of multilateral desirability? There are some obvious steps whereby we could resolve both of these problems, and it is worth looking at these before going on to ask about whether we in Erewhon would be likely to take those steps.

The unity problem arises from the fact that different, if overlapping, sets of desiderata support that which is desirable from the individual, the joint, and the common perspectives. But there is a salient way in which we might hope to resolve this, which is to put those desiderata together and determine what they support in aggregate.

Many desiderata are going to be involved in a number of practical perspectives, as when a property, P, that makes something robustly attractive within my individual point of view also makes it robustly attractive within one or another joint viewpoint or within the common viewpoint of an unbounded group. That means that a desideratum like P can weigh against one set of desiderata within one viewpoint, against others in another, in determining perspective-specific desirability. And this shows that there is no problem in principle with thinking that desiderata that figure in different viewpoints may weigh against one another in determining what is desirable from a viewpoint that transcends particular perspectives: that is, what is personally most desirable overall.

The desiderata that weigh against one another within any particular viewpoint may not uniquely determine a desirable alternative from that perspective but will presumably do so in many cases. Similarly, the desiderata from different viewpoints that weigh against one another in determining desirability of a kind that transcends viewpoints may not always identify a single alternative as the one that is personally most desirable overall. But there is no reason to think that

they will not be able to do this in the majority, perhaps the vast majority, of cases.

 (p.184) If any one of us can aggregate different desiderata to determine what in many cases is desirable overall by our lights, then we can achieve unity in our judgments of desirability. Consistently with this resolution of the disunity problem, however, I might give a totally different extension to the emerging concept of desirability from that which you or others give it. I might let the desiderata aggregate to support one set of options in a range of relevant choices; you might let them aggregate to support quite a different set. In that case, the concept of such unified desirability would be idiosyncratic to each of us: in effect, I would work with a unilateral concept of desirable$_{me}$ and you with a unilateral concept of desirable$_{you}$.

While the move described might get rid of disunity, then, it would not provide the means of overcoming the lack of universality and delivering a fully multilateral concept of desirability. Indeed, as we saw, getting rid of the disunity, and developing the concept of what is personally most desirable for someone in a given situation, might make the need for a multilateral concept of desirability even more pressing. We might feel unsure about being able to rely on others to act congenially if they are moved by what is personally most desirable for them. And in any case, we might find it difficult to determine what is personally most desirable for this or that individual; that might require an exploration of their personal psychology.

As the disunity problem is resoluble with the means at our disposal in Erewhon, however, so the non-universality problem ought to be resoluble, too. We will be able to resolve it by restricting the desiderata relevant to determining what is multilaterally desirable to properties such that we can welcome the prospect that they should move others as well as ourselves. Where the resolution of the first problem would require us to look in aggregate at the desiderata mobilized under different practices, the resolution of the second would require us to filter the desiderata, putting aside those that put people in inescapable competition with one another.

The desiderata excluded under this filter may include both of two kinds of properties, which we may describe respectively as rancorous and rivalrous attractors. Rancorous attractors are ones that make **(p.185)** something attractive to me or to my group, but only at the cost of imposing harm on other individuals or groups. They would make an action attractive despite its harming others: say, despite its meaning that we hurt them physically, seize some of their property, or threaten their lives. Rivalrous properties are ones such that as a result of scarcity, the fact that some individuals or groups enjoy them in their own case necessarily means that others cannot enjoy them in theirs. These

properties would make an action attractive for its giving me an advantage over others, although it would not require us to harm them directly.

Thus, to illustrate rivalrous properties, we may each find esteem or power or wealth attractive from our particular individual or group viewpoints—we may see it as an agent-relative or group-relative good—despite the fact that the more we enjoy such a benefit, the less others can do so. Everyone can seek to be famous or influential or to be better off than others, of course, and it may be a non-rivalrous attractor that they should do so. But not everyone can win that prize, since as a matter of logic not everyone can be famous or influential, and not everyone can have more goods than others. And so, it cannot be a non-rivalrous attractor in a certain arrangement that it would deliver such a benefit for me or for any other individual.[10]

What are the non-rancorous, non-rivalrous desiderata that would pass the filter for multilateral desirability? They are certainly going to include the agent-neutral, group-neutral desiderata that would make a scenario attractive from no matter what perspective or standpoint: desiderata like truth-telling or peace or prosperity. But they can also include agent-relative or group-relative desiderata that count as concordant or non-competing.

These are relativized properties such that it is possible for each of the relevant agents or groups to instantiate them at once, and to do so without necessarily reducing the level of satisfaction in other agents or groups. Plausibly, they may include the property of my children doing **(p.186)** well, my keeping my pledges, some group to which I belong prospering, my actively pursuing such an agent-relative or group-relative result, and even my pursuing some rivalrous prize in an open competition.

In a term from R.M.Hare (1952, 1981), all such concordant agent-relative desiderata, like all agent-neutral desiderata, are universalizable properties. They are such that if we prescribe realizing or pursuing them for ourselves, then without any inconsistency we can prescribe their realization or pursuit for anyone in a similar position to ours. Thus, we can universally prescribe in a suitable context that each of us should pursue or realize a certain agent-neutral good. And, despite the agent-relativity involved, we can universally prescribe that each of us should look after our own children, keep the pledges we make, look after particular groups to which we belong, pursue those ends as best we can, or pursue even a rivalrous end in a suitably open competition.[11]

If we were able to develop a concept of the multilaterally desirable, following steps like these, would it be likely to make various options substantively desirable? Would it be likely to make them desirable enough on every side for each of us to expect others to be moved by the fact that an option has that feature? Assuming, in line with our narrative, that we in Erewhon manifestly

stand to benefit from being able to rely on others and to get others to rely on us —assuming in that sense that we are invested in relationships of mutual reliance —it is very likely that we will each be responsive to what we see as the multilaterally desirable. Indeed it is likely that any option that is multilaterally desirable will generally prove to be also the personally most desirable option for us to take.

(p.187) These observations show that a concept of the multilaterally desirable is bound to be available, at least in principle, within Erewhon. It may be that we depend on practices of avowal and co-avowal to access concepts of the individually, commonly, and jointly desirable. And it may be that none of those concepts has the status of a master category, as the concept of the commonly credible has that status in the case of belief; it may be that they leave us each short of a unified or universal concept of desirability. But that would still allow us, in the case of any choice, to aggregate and unify the desiderata relevant to different modes of desirability, to filter out those desiderata that make universality inaccessible, and to let the resulting set determine the multilaterally desirable alternative.

Why we would repair the shortfall

Is there any reason to think that we would introduce the category of the multilaterally desirable in Erewhon and thereby overcome the problem of non-universality? Is there reason to think that that category would emerge more or less spontaneously among us? Or is there reason, at the least, to believe that we would be motivated as a community to invoke it as a solution to a collective problem: to contract, as it were, into such a resolution?

There is certainly reason why we might be motivated as a community to introduce the category, since doing so would establish a firm prospect of overcoming various conflicts over desirability, as the master category of common credibility entrenches a prospect of overcoming conflicts on issues of credibility. It would point up the problems that the absence of that category creates among us and motivate a search for the more or less obvious sort of solution just sketched. And it would do so in the presence of conceptual resources that made the required sort of contract accessible.[12]

(p.188) But there is no need to invoke such a collective move in order to make sense of why we might get to think in terms of the multilaterally desirable. For there is a fairly plausible story to tell about how we might generate the category spontaneously: how we might generate it, first, by adjusting to the challenge of disunity and generating in response a unified concept of the personally most desirable; and then, by adjusting to the problem of non-universality, generating in the process a concept of what is multilaterally rather than just personally desirable.

We can easily imagine, taking the disunity problem first, that you will naturally put pressure on me to say where I stand when the different options I face appear to score differently in different modes of desirability. Suppose that it seems to be individually desirable for me to take one option, jointly desirable for me to take another, and commonly desirable for me to take a third. Suppose, indeed, that I myself openly acknowledge this situation, as I will presumably have to do. In such a situation, you will be likely to despair of determining where I stand if I do not go beyond those particular perspectives to say what is desirable for me in a more outright fashion. And since the likelihood of your despairing of me will be clearly repugnant from my point of view, it must be a matter of manifest expectation that I will come off the fence in any such scenario.

What holds for me in this sort of situation will hold equally for you or anyone else. There will be enormous pressure on each of us, therefore, to acknowledge that not only is there an issue as to which of a number of scenarios is most desirable in the individual or joint or common mode; there is also a question as to which is most desirable in a manner that transcends those modes. But if I or you or anyone else responds to that pressure, we will each make room for a concept of what is overall desirable for us, regardless of practical perspective. And, the pressure and the response being manifest, we will do so as a matter of common awareness across the community.

With access to the concept of what is personally most desirable for each of us, the problem of non-universality will be particularly pressing. For we may begin to feel that we do not have sure ground for relying on someone else to act in a congenial way, even someone who has made an **(p.189)** explicit commitment to us, unless we can see that acting in that way is personally the most desirable option available to them. And we will not be in a ready position to determine that one or another option satisfies this condition: that would seem to require a customized sense of their particular psychology.

There are some cases, of course, where we might find our way beyond this problem of non-universality. These are cases in which the alternative that is desirable$_{me}$ is manifestly also desirable$_{you}$. This convergence would be bound to attract notice, leading us to cast the option as desirable from a shared, interpersonal perspective—desirable$_{us}$—and inviting the thought that it may even be desirable$_{all}$. Any option that was desirable$_{all}$, of course, would count as multilaterally desirable: it would satisfy all relevant desiderata for each, including all that are non-rancorous and non-rivalrous, and would be bound to prove desirable on all sides.

Of course, cases where the notion of the desirable$_{us}$ clearly applies—and, potentially, the notion of the desirable$_{all}$—are bound to exceptional. But the experience of such cases in Erewhon is bound to underline for us the attraction

and the possibility of finding an interpersonal perspective, and achieving convergence, in other cases too.

In those cases, achieving full convergence will be problematic just insofar as we are each exposed to the effect of rancorous or rivalrous desiderata, and are liable to seek ends that require directly harming others or to pursue essentially competitive goods for ourselves like esteem or power or wealth. But that means that the search for an interpersonal, convergent perspective will naturally lead us to focus on that which is attractive for each of us on a non-rancorous, non-rivalrous basis. That basis ought to be enough to secure the universal attraction of various options, given that we are invested in building relations of mutual reliance with one another.

Thus, the search for a convergent, universal perspective would be likely to lead us, by this account, to make use of a notion of multilateral desirability that applies to anything that is supported on the basis of non-rancorious, non-rivalrous desiderata. And that notion would be a salient and appealing candidate for playing in most cases the role that the concept of the desirable$_{us}$ or the desirable$_{all}$ can play in some. The **(p.190)** concept would be salient insofar as such relationships of mutual reliance could clearly prosper only if we suppressed the attraction of rancorous and rivalrous properties. And it would appeal to each of us insofar as we could plausibly invoke it to support our disposition to behave in a manner congenial to others—say, to live up to our commitments—and to persuade others of our reliability on that front; the support provided would be persuasive in virtue of our manifest investment in relationships of mutual reliance.

If this line of thought is sound, then the sorts of pressures that lead us in Erewhon to make use of avowals and pledges in communicating our attitudes, and the sorts of pressures that lead us therefore to think in terms of the credible and the desirable, are likely to provide us with prompts that push us eventually into invoking a concept of multilateral desirability. The concept will be a resource that is going to be more or less manifestly available to us and that we can use to the manifestly attractive effect of assuring others of our broadly congenial dispositions. Thus, the trajectory that the narrative has traced leads us plausibly, if not inexorably, towards that concept.

Assuming that concept gets established in the community, it will enable us to form beliefs in common about what is multilaterally desirable in this situation or that, and about what options are multilaterally desirable. And, of course, it will also enable us to avow and presume to co-avow a belief that an option is multilaterally desirable, whether by expressing the belief in an assertion that it is, or by ascribing the belief to ourselves, or by explaining why it is true. Avowal and co-avowal are recursive operations, as already emphasized, so that the fact that the concept of the multilaterally desirable would not be available in the

absence of such practices is quite consistent with our being able to rerun the practices with beliefs that presuppose the availability of that concept.

The range of the multilaterally desirable

How substantive is the agreement we are likely to achieve about the multilaterally desirable option in any choice: say, in a choice between **(p.191)** alternative actions that an agent might take or between alternative arrangements that we as a community might establish? Agreement will be easily achieved in cases in which one option in the choice satisfies all the non-rancorous, non-rivalrous desiderata that are satisfied by others and satisfies them in a higher measure; or, satisfying some of those desiderata in at least equal measure, satisfies others better. And it will be relatively easily achieved in other cases to the extent to which the weightings we attach to the different relevant desiderata are in more or less the same range.

The relevant desiderata that we recognize in certain choices may be weighted differently, of course, and may block the emergence of full-scale convergence. In many cases, this will generate a range of different views but still allow continuing discussion that is aimed at clarifying the issues involved and perhaps reducing differences of weighting. In other cases, however, the weightings may be so divergent that, no matter how much clarification is achieved, divergence is going to prove unavoidable. In those cases, it will be indeterminate whether this or that option in a choice is multilaterally desirable.

Notwithstanding such failures of convergence on the extension of multilateral desirability, however, we may expect to achieve actual agreement over a broad range of issues, even if we are likely to leave some questions unresolved or treat them as irresoluble. Thus, it is likely that we in Erewhon will agree about the multilateral desirability of many types of choice, or at least about the desirability of most instances of those types.

Consider choices of the kind that are generally resolved by social norms of the kind introduced in the second chapter. Each of us is likely to be sensitive to the desiderata relevant to the multilateral desirability of conforming to such a norm. We are likely to agree that it is multilaterally desirable to tell the truth, abstain from violence, avoid fraud, and so on. Those norms will represent standards such that, as a matter of common awareness, we all take conformity to be multilaterally desirable, at least in most contexts: specifically, in contexts where their demands are clear and conformity is not likely to trigger any exceptional, multilaterally undesirable costs.

(p.192) Standards that are shared in this sense may be supererogatory: they may require a capacity for self-sacrifice—a degree of heroic virtue—that few of us will be expected to possess. While we may well share some supererogatory standards of multilateral desirability in Erewhon, we are also likely to share

standards that we take people generally to be able to meet. These will include standards like those that preclude lying, violence and fraud, which are certainly within our normal capacity to follow—otherwise they could not get established under reputational pressures—and within our normal capacity as a matter of common belief. They will count in that sense, not only as shared, but as routine standards of multilateral desirability.

When standards of multilateral desirability are shared and routine in this sense, they will constitute norms of desirability, as the notion of a norm was defined in chapter 2. Being shared and routine, people will generally conform to what they present as multilaterally desirable patterns; conforming to those patterns will generally be expected to have reputational benefits, therefore, not conforming to have reputational costs; this expectation will help to support the general pattern of conformity; and all of this is likely to be a matter of common awareness. This is to say that many of the purely social norms mentioned in chapter 2 are likely to double as norms of multilateral desirability in Erewhon: that is, as norms such that following them requires understanding what it is to be desirable in that sense.

The discussion in chapter 2 showed how pre-social norms that are supported by reputational forces, without our necessarily recognizing them or tracking them, will become properly social norms once they satisfy a common awareness condition. At that point, we will see meeting their requirements as a condition for full acceptance in the community and can invoke that condition as a desideratum of complying. When social norms double as norms of multilateral desirability, the difference made is that we can now see meeting their requirements, not just as a condition for social acceptance, but as a condition for satisfying the demands of multilateral desirability. The reputational forces will remain in place as an engine motivating conformity, as will the social desideratum of conformity. But the norms will gain a new aspect under which **(p.193)** conformity to them appeals: one that is provided by the idea of the multilaterally desirable.

There are also likely to be norms of multilateral desirability established in Erewhon that reflect practices of avowal and pledging that had not yet appeared in the narrative of chapter 2. Given the importance assigned on all sides to these practices, and given their central role in social life, we are all likely to regard it as multilaterally desirable that people should live up to the unquestioned demands of their avowals and pledges: that they should display fidelity to their commitments. And equally, we are certainly likely to regard such fidelity as lying within people's normal capacity to follow. Standards of fidelity will be both shared and routine in the society, and will constitute further norms of multilateral desirability.[13]

The norms of multilateral desirability that prevail at any one time in Erewhon may not coincide with those that prevail in another society or in Erewhon at another time. It is very likely that conversation about matters of multilateral desirability may affect the standards that the members of any society are likely to endorse in a shared manner at any time. And it is equally likely that as conformity to any such pattern rises, the disesteem for not conforming will increase, and conformity may become motivationally more accessible, with the result that an aspirational ideal may turn into a routine standard. The assumption made about Erewhon is that there will almost certainly be some standards of multilateral desirability that are shared and routine at any period, not that any particular set of standards will achieve this status or that they will remain in place for all time.

Multilateral and moral desirability

These observations argue that as we in Erewhon would each come to access a range of practice-specific concepts of desirability, so in all likelihood we would evolve a unified, universalized concept of multilateral **(p.194)** desirability. This would give us a master concept of desirability to rival the master concept of common credibility and would ease the path to mutual reliance. The argument marks a crucial development in the narrative, for the concept of the multilaterally desirable coincides broadly with the concept of the morally desirable, as that was outlined earlier.

Thus, the concept invokes considerations that are relatively unrestricted in the range of interests invoked, relatively unrestricted in the standpoint adopted, and consequently fit to play an authoritative role in adjudicating certain clashes among other judgments of desirability. The concept of multilateral desirability is unrestricted in the range of interests that it reflects since it is designed not to reflect just the particular interests of one individual or grouping. It is also relatively unrestricted in the standpoint adopted, since it is designed to unify the perspectives offered by different practices for any individual, and to universalize the perspectives of different individuals. And being designed to adjudicate conflicts between the rival interests and investments of different people, it is bound to have the authoritative cast associated with the concept of moral desirability.

The notion of the multilaterally desirable not only connects with morality in this manner; it also converges with the concept of moral desirability in more detailed ways. It satisfies the three generic constraints satisfied by any concept of desirability, including the moral. And it conforms to the two specific constraints associated with the concept of moral desirability in particular.

The generic constraints are those of grounding, divergence, and governance. If one option in a choice counts as multilaterally desirable and others not, then there must be a difference in the desiderata that ground their relative

desirability. If an option is desirable in that sense, or even if I judge it to be desirable in that sense, it may still be that what I desire diverges from that judgment as a result of a distraction or disturbance that I fail to resist: say, the disturbance generated by a rancorous or **(p.195)** rivalrous attractor. And my functioning properly as an agent among agents—as someone invested in relationships of mutual reliance—requires that the judgment of multilateral desirability should dictate or govern what I actually desire in relevant contexts. This will be so, at any rate, when other things are equal: for example, when the multilaterally desirable is not beyond my capacity and, to anticipate discussion in the next chapter, I am fit to be held responsible for pursuing it.

The first of the specific constraints on the concept of the morally desirable requires that you and I should have the same content in mind when we judge that it is desirable for anyone, whether anyone in general or anyone in a certain position, to choose a given option. And the second requires that it should be true or false that the option is desirable—assuming the issue is determinate—so that there is no possibility that it might be true by your criteria as an assessor, false by mine.

Is the concept of the multilaterally desirable likely to meet these constraints? It must do so if it is to get us out of the problem raised by rival personal standpoints, satisfying an important part of its design specification. If it is to facilitate interpersonal understanding and reliance in the required manner, then it must rule out both the relativity of content that the first constraint forbids and the relativity of truth-value that the second constraint outlaws. It must direct us to a range of issues that we may hope to explore and resolve in common.

These considerations argue that the concept of multilateral desirability, which would be likely to evolve among us under the pressures described, can be identified with the familiar concept of moral desirability. In at least this respect, then, we in Erewhon would be more or less bound to develop an ethical way of thinking and prove ourselves an ethical species.

The convergence of the idea of multilateral desirability on the idea of moral desirability is hardly surprising, at least on many received accounts of what it is to take the moral point of view. David Hume gave an account of this point of view that makes a nice connection with the multilateral idea (Sayre-McCord 1994). Asking what someone does in adopting the language of morals and using it to assess others, Hume (1983, s9.6) answers: "he expresses sentiments, in which, he expects, **(p.196)** all his audience are to concur with him. He must here, therefore, depart from his private and particular situation, and must chuse a point of view, common to him with others: He must move some universal principle of the human frame, and touch a string, to which all mankind have an accord and symphony."[14]

The developments charted in this chapter do not yet give us the concept of moral obligation, which figures even more prominently in ethics than does that of moral desirability. But if we come in addition to develop the idea of responsibility, we are also going to be in a position to introduce a concept that plays the role of the obligatory, under the construal adopted here. It will be morally obligatory for someone to choose a certain option under that construal if it meets two conditions: it figures among those options that fall within the domain where the agent is fit to be held responsible, and it is morally the most desirable of those options. This means that we will have access to the concept of moral obligation—and to the related concepts of moral prohibition and permission—if it is possible for us to gain access to the notion of responsibility. And that possibility is the topic of the next chapter.

Notes:

(1.) An option is a possibility that will be realized, given how things are in the world, depending on what you want (Pettit 2018a). This means that an option will be disjunctive in the fashion of X-or-Y in any case, like that envisaged in the text, where you rank the disjuncts equally and are happy to allow chance, or some such arrangement, to select between them. What of the case in which you can opt to do something that will result in X or in Y, depending on chance, but you cannot opt for X or for Y independently? In that case, X and Y will count as possible outcomes of the basic option, and that option is likely to be identifiable other than just in disjunctive terms as X-or-Y.

(2.) The first claim is likely to be widely accepted, since there are many sets of alternatives where moral considerations or reasons are irrelevant in ranking the members, so that the member that is overall most desirable may not be morally desirable. The second claim is at the center of a controversy about whether, as it is put, there could be most reason overall to do something that runs counter to moral reasons. For rival responses, see Portmore (2014), who answers in the negative, and Dorsey (2017), who answers in the positive.

(3.) This, by a plausible analysis, is what happens when Huck Finn judges that he ought to report Jim, the runaway slave, but does not do so: this, presumptively, because of actually believing that it is desirable not to report him (see Arpaly 2003; Joshi 2016).

(4.) There is an important issue as to how such prescriptive principles relate to constraints on credence and utility supported in one or another form of decision theory. For some consideration, see Pettit (2016b).

(5.) This is to say that the ideal self may advise that the actual self should behave in a manner that takes account of difficulties the ideal self does not itself have to deal with. On this lesson, which also applies in the idealizations considered later, see Smith (1994).

(6.) Of course, it may be that the data do not require any particular belief but allow both of our beliefs. For simplicity, as mentioned earlier, I concentrate here on the more straightforward case.

(7.) This is not to deny that there are some standpoints, say those associated with certain forms of oppression, such that it may take enormous efforts of empathy on the part of outsiders—and a willingness to trust the testimony of those occupying such a standpoint—to grasp what is revealed therein. Similar points apply, of course, in the case of the commonly desirable. See Jones (1999) and, for a general perspective, Fricker (2007).

(8.) If that new subject is to be conversable itself, of course, operating like one of us, then it must presumably submit to the discipline of common credibility. Otherwise it could not present as a subject that others can understand and rely on (Buchak and Pettit 2014). The difficult issues that this raises, however, are beyond the reach of the current discussion. One case in which they arise is when a corporate group seeks to develop group judgments of desirability, including multilateral or moral desirability; although that case is not considered here, later discussion indicates that it is possible.

(9.) The treatment here ignores the fact that it may prove to be desirable, in whatever mode, to hold a belief in what is not commonly credible. We ignore such practical or state-centered reasons, as distinct from theoretical or object-centered reasons, for holding by certain beliefs (Parfit 2001).

(10.) We might prescribe that people should each seek fame or influence or a greater share of some goods than others—this, because we think that the competition would have a result that is commonly desirable—while fully recognizing that they cannot all attain that prize. The property of seeking the prize is non-rivalrous, the property of winning the prize is rivalrous. The distinction is similar to that drawn by Derek Parfit (1984, 53) between prescriptive theories that are 'directly collectively self-defeating' but not 'indirectly collectively self-defeating'.

(11.) Hare himself fails to recognize that certain agent-relative desiderata can pass his universalizability test, so that he takes it to support the view that only the commonly desirable can play a role in determining what he thinks of as moral desirability. He is in that sense a classical consquentialist.

(12.) Invoking a contract at this point would not presuppose the sort of desirability concept it seeks to explain; it would differ in that regard from the contractual stories that were contrasted in chapter 1 with stories of unplanned emergence.

(13.) This illustrates yet another aspect of the recursion stressed at various points. We may depend in Erewhon on the practice and concept of avowing and pledging in order to develop the concept of multilateral desirability. But that does not prevent us from applying that very concept to those practices.

(14.) I am grateful to Ralph Wedgwood for drawing my attention to this Humean passage.

Access brought to you by: