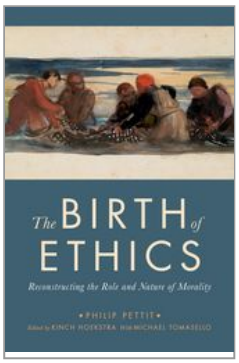

The Birth of Ethics: Reconstructing the Role and Nature of Morality
Philip Pettit and Kinch Hoekstra

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

Reconstructing Morality

Philip Pettit

DOI:10.1093/oso/9780190904913.003.0003

Abstract and Keywords

Ethics requires people not just to be moved by relatively altruistic feelings to perform relatively altruisic actions, but to be moved in this way by considerations that they conceptualize in ethical terms or concepts. Those concepts come in many forms, but two important families cluster around, first, the idea of desirable options and, second, the idea of agents who are fit to be held responsible for taking or not taking such options. The aim of this book is to explain the emergence of ethical concepts and practices in a naturalistic manner that vindicates realism. Such a story of emergence would help to make sense of ethics, directing us to the sorts of properties predicated in talk of desirability and responsibility. But in order to do so, it would have to start from a naturalistically intelligible, pre-moral starting point—ground zero—and explain in naturalistic terms how people in that society would be likely to make a cascading series of adjustments that would eventually lead them into ethical space. The project of developing such a story is akin to various approaches taken in other branches of philosophy, embodying a conceptual genealogy, and employing something like the method of creature-construction, but has not been undertaken before for ethics, at least not in the way it is undertaken here.

Keywords: Naturalistic, realism, desirability, responsibility, emergence, ethics, altruistic, genealogy, creature-construction

The aim of this project is to offer an account of ethics or morality—I use the terms here as synonyms—that makes sense of how we come to be an ethical species. This opening chapter offers an account of the reconstructive approach

taken to that project. But before developing that account, it will be useful to say a little about what is involved, as I see things, in our being an ethical species.

Being an ethical species

We human beings are ethical creatures insofar as we are affectively and behaviorally responsive to ethical or moral considerations—that is, considerations mobilized by moral concepts—and are disposed, however fallibly, to act as they dictate. The motives and actions that ethical considerations evoke will be generally altruistic, going against the grain of more self-serving inclinations. But even altruistic emotions and actions will not count as moral, unless the considerations that we recognize as supporting them include many of a moral character. The altruistic responses of other animals toward their young need hardly count as moral responses. And that, presumably, is because considerations articulated in moral terms play no part in prompting or in regulating those responses.

The concepts that mobilize ethical considerations come in many varieties. They include concepts of what you ought to do, of the treatment you owe me or someone else, and of what therefore is our due and your duty. Thus, they include concepts of how it would be good or bad for you to behave as an agent—say, of what you are forbidden, required, **(p.14)** or permitted to do—as well as concepts of what I, qua beneficiary, can claim as a matter of right or justice or desert: this might be a claim that you should keep your promises or show me a modicum of respect. And, moving from an ex ante to an ex post perspective, ethical concepts also include ideas of censure and commendation, guilt and shame and pride.

Ethics requires that we interact with one another as these concepts dictate that we should interact, but also that we act in that way, at least in some part, because of the considerations that the concepts put within our reach. We must be ready and able to apply the concepts in ethical practices like those of praising and criticizing various types of behavior; granting claims to others and making claims upon them; exhorting ourselves and exhorting others to act appropriately; and censuring or commending one another for how we act. And those exercises must not be irrelevant to how we behave. Even if we are each spontaneously disposed to live up to accepted moral standards, we must be ready to activate the practices, whether in dealing with ourselves or others, in the event of the disposition failing.

Desirability and responsibility

While ethical concepts vary, however, and while they can mobilize a wide range of distinct considerations, they all serve to mark different grounds on which actions count as morally desirable, on the one side, and agents count as fit to be held morally responsible, on the other. This makes it possible to use the concepts of desirability and responsibility to explicate other ethical concepts, or at least to

identify the role they play. In order to keep things simple and tractable, this study will concentrate on those two dimensions of morality and reconnect with other concepts only in the final chapter.

This focus does not tilt things in favor of any particular understanding of morality; or at least, as we shall see in that final chapter, not in an independently indefensible manner. Thus, it does not beg the question between rival moral theories like consequentialism and non-consequentialism nor between approaches that differ, as we shall see, on **(p.15)** the relative priority of desirability and responsibility. The explanation of ethics defended makes sense of why people should be drawn to one or the other side in such debates but does not provide grounds on which to choose between the sides.

The notions of desirability and responsibility are tightly connected with one another. When you judge that one option in a given choice is morally more desirable than alternatives, then, other things being equal, you explicitly or implicitly recommend or enjoin or prescribe its performance by the relevant agent, be that yourself or another. And when you hold that the agent is fit to be held responsible for the choice, you maintain that other things are indeed equal: the agent has the capacities and satisfies the other conditions required for such fitness. In that case, then, you prescribe the option without reserve or, depending on what was actually chosen, you deem the choice praiseworthy or blameworthy in retrospect: you commend or condemn the action and, if you are the agent, you feel pride or guilt about how you behaved.

Moral desirability

Any concept of desirability, including that of moral desirability, is designed to mediate prescriptions for what we should desire—including, by implication, intend—and in that sense, it is essentially practical. It will mediate suitable prescriptions in relation to agents or agencies who have the capacities required for being fit to be held responsible for the choice made. Thus, any such concept can be used to prescribe for what we should individually do, for how we should individually be, or for what we should collectively establish. It may prescribe for how we should individually make things be in the external world or for what sorts of internal attitudes or states we should cultivate or for how we should coordinate with others.

Prescriptive concepts come in many varieties, of course, and desirability in particular may mean moral desirability or desirability of some other kind: for example, desirability in law or etiquette, epistemological desirability, or desirability as a matter of prudence or patriotism. And, **(p.16)** as desirability may assume different forms, so too may responsibility. This may mean, not responsibility for living up to moral demands, but responsibility for living up to the demands of law or etiquette, the epistemological demands of evidence, or the demands of prudence or patriotism.

What distinguishes the concept of moral desirability from such other concepts of desirability? There is no theory-independent characterization that promises to pass muster on all sides, but there are some general orientating remarks that ought to be acceptable on most.

Whenever something is desirable, by all accounts, it is desirable in virtue of having certain properties or, equivalently, in virtue of satisfying certain considerations. And by almost all accounts, the considerations that make an alternative morally desirable, rather than desirable in any other way, are characteristically unrestricted or comprehensive.

Those considerations are unrestricted in two respects. First, unlike considerations of prudence or patriotism, they are not restricted in range to the self-interests of a particular individual or country or whatever. If they argue for the moral desirability of someone's taking a prudent option, for example, they will do so in light of the impact on others as well. Second, the considerations are not restricted in standpoint to considerations about what a code of law or etiquette requires; to considerations about what, in certain cases—for example, cases of belief—the available evidence supports; or to considerations about what promotes a set of projects, perhaps altruistic in character, espoused by some individual or group. If moral considerations converge with any such considerations in supporting an option, they will do so in light of the overall impact of obeying the code, sticking with the evidence, or supporting the projects.

Different theories of ethics or morality are liable to differ on what more can be said about the considerations that are relevant to determining moral desirability; consequentialist and non-consequentialist theories, as we shall see in chapter 7, differ in this way. But one assumption common to many different viewpoints is that, as a result of the relative lack of restriction in range and standpoint, what these considerations support—or at least what they support for agents and agencies that are fit to be held responsible—they support in a comparatively weighty or **(p.17)** authoritative manner. The considerations that argue for the moral desirability of one or another option are designed to be trump cards that outweigh competing considerations that reflect just a restricted range of interests or a restricted standpoint of concern.

These remarks should be enough to provide us with orientation for the moment. We return to the issue of how to think about desirability, and moral desirability in particular, at the beginning of chapter 5.

The project in this book is to make sense of how we could have come to be capable of registering matters of moral desirability and responsibility, capable of holding one another to moral account, and capable of regulating ourselves so as to prove accountable. In the three sections of this chapter, we look in turn at the

genealogical or reconstructive methodology to be followed in the project; at the different constraints that it must satisfy; and at some precedents for its employment. Readers less interested in methodology may wish to go straight to Chapter 2.

1.1 Reconstructing ethics naturalistically

The naturalistic challenge

Because of being inherently prescriptive, ethics raises a problem for those of us who embrace naturalism, holding that the world we live in is an austere place that conforms to the image projected in natural science. Naturalists maintain that all the properties realized in the actual world are naturalistic in the following broad sense: they are of a kind with the fundamental properties that are liable to figure in natural science—properties like mass and charge and spin—or with properties that are actually realized by one or another configuration of such properties; the latter might be illustrated by properties like being dense or impenetrable, audible or visible, asymmetrical or pear-shaped.[1] And naturalists **(p.18)** hold in addition that everything that happens in the actual world does so under the force of the laws, deterministic or in-deterministic, that operate at the most fundamental, presumably sub-atomic level; no new forces—that is, no new fundamental forces—appear at higher levels of realization.

The fundamental properties recognized in natural science—even natural science extended by mathematics—do not include prescriptive properties like desirability and responsibility. And so, there is a serious question for naturalists as to whether prescriptive properties—and, in particular, the properties of moral desirability and responsibility—really are features of the universe we confront. They will presumably enjoy that metaphysical status, on a naturalistic view, only if they can be realized by one or another configuration of fundamental properties. But how could they be realized by properties that have no connection themselves with prescription? That is the naturalistic challenge.[2]

The assumption of realism

Naturalistic philosophers have sometimes responded to this problem by invoking the possibility of projection or illusion, seeking to debunk the idea that, by naturalistic lights, desirability and responsibility are bona fide properties. They have opted for representing ethical talk as fundamentally emotive or expressive, for example, rather than taking it to be descriptive of any features in the world (Stevenson 1944; Ayer 1982; **(p.19)** Blackburn 1984; Gibbard 1990).[3] Or they have held that in speaking ethically we treat desirability and responsibility as if they were real properties when actually they are not: consciously or otherwise, so this story goes, we regard them as fictions (Mackie 1977; Joyce 2006).

The idea in this book is to resist downgrading ethical discourse in any such manner and, without forsaking naturalism, to try in the spirit of realism to vindicate the assumption that there really are properties like desirability and

responsibility in the world and that they have an impact on our actions. These properties do not belong with fundamental, naturalistic properties like mass and charge and spin that figure explicitly in the basic sciences. But the assumption is that despite being non-prescriptive themselves, those fundamental properties can realize or ground the properties we discern in identifying desirable actions and responsible agents. The fundamental properties can realize such properties, so the idea goes, in the way in which the pixels on a television screen realize the patterns or properties that we register in following any TV program. They ground them in the sense of both guaranteeing and explaining their presence.[4]

The pixels on a television screen differ from the patterns they ground insofar as they are not anthropocentric—they can be defined in electronic terms that make no reference to human beings—whereas the supported patterns are: it is only because of our human interests that the smile on a presenter's face or the handshake of the protagonists in a drama are patterns worth noticing. To hold that as the television patterns are grounded in the pixels, so the properties of desirability and responsibility are grounded in equally naturalistic properties, is to allow that desirability and responsibility may be anthropocentric in a similar fashion. For all that naturalistic realism requires, these properties may **(p.20)** be patterns that deserve notice only from the standpoint of creatures with distinctively human sensibilities, engaged in characteristically human practices.

The standard naturalistic strategy

The standard way to vindicate a naturalistic realism about ethical properties is to try to provide a naturalistic reduction for them. In a familiar, relatively strict form, this reduction would involve two claims. First, that what it takes for an ethical property to be instantiated is that this or that set of non-prescriptive conditions are satisfied. And second, that those conditions can be satisfied in the actual world by one or another configuration of broadly naturalistic properties: that is, scientifically fundamental properties or properties that are realized by one or another configuration of fundamental properties.

The first claim is an analysis of the conditions that make it appropriate to apply the ethical term or concept, by our intuitive understanding; this is usually developed by the method of cases, which involves thinking about the range of scenarios where we would apply the concept, and the range where we wouldn't.[5] The second claim is an empirical thesis to the effect that those conditions can be satisfied in this or that naturalistic configuration.[6]

The two-claim pattern exhibited by such a reductive analysis may be illustrated in the naturalistic reduction of a property like that of believing that p. This might hold, to gesture loosely at a functional style of analysis, that the belief is instantiated in a subject in virtue of a more or less reliable disposition to perform actions and adjust attitudes as if **(p.21)** it were the case that p.[7] And it would maintain, as a matter of empirical grounding, that in the actual world—

although perhaps not in every possible world—this disposition consists in the subject's having a suitable naturalistic character: having a neural or electronic configuration—maybe this, maybe that—that generates the required pattern of action or adjustment.

Not all naturalists will agree that this is the way to reduce a property like believing that p. Some may ignore the analytical thesis altogether, and claim only that every instance of the property is just identical with the instance of a naturalistic property (Block and Stalnaker 2000). Others may keep the analytical thesis in a weaker form, treating it as a proposal about how best to understand the property to be reduced, not necessarily as an analysis of how the property is conceptualized in ordinary usage. But the strict form of reductive analysis will serve us well as a foil to the alternative introduced here.[8]

For reasons of convenience, however, this presentation of the strict form of reductive analysis is simplified in one regard. It assumes that there is no difficulty in identifying the precise conditions that make it appropriate to apply the target concept, and to ascribe the target property; and that there is no difficulty, therefore, in identifying the naturalistic configuration of properties that grounds the instantiation of the property. But in practice, there will often be a choice available as to which set of conditions to privilege analytically and which configuration of grounding properties to take as grounding the property to be reduced (Johnston 1992, 221–222). The reductive analysis of causation will assume a different form, for example, depending on whether or not we assume that every cause must take time to have an effect or that it must be connected by intervening factors with an effect.

(p.22) What set of conditions to privilege, and what naturalistic configuration to take as ground of the property, may be determined in part by external considerations. We may make that determination on the basis that one candidate does better than others in serving the role that the concept plays in ordinary exchanges; in supporting our general theory of the domain in question; or even in making the naturalistic grounding of the property more straightforward (Burgess and Plunkett 2013).[9]

The reconstructive alternative

The alternative approach does not begin with an analysis of desirability or responsibility and then seek to argue that conditions sufficient to realize such properties are satisfied naturalistically in the actual world. It starts rather from a naturalistic story about how recognizably ethical terms and concepts could have emerged among creatures of our ilk and could have played a referential, yet prescriptive role in registering bona fide properties of the world. And then it argues on that basis for a naturalistic realism about desirability and responsibility.

The argument involves two claims. First, that insofar as the terms or concepts that emerge in the story respond to the same sorts of prompts, and serve the same sorts of purposes, as our actual ethical terms, the properties they predicate are good candidates for the properties we ourselves predicate with our terms. And second, that since the appearance of those concepts in a predicative role is naturalistically explicable, the properties they ascribe—and the properties ascribed by our counterpart concepts—must be naturalistic, too; if the concepts ascribed non-natural properties, after all, then those properties would presumably have played a role in explaining how the concepts came into use.

(p.23) These two claims correspond broadly to the two claims in a strict reductive analysis. The first replaces the analytical claim with a thesis to the effect that ethical concepts are expressively equivalent to the concepts introduced in the story. And the second replaces the empirical claim with a thesis to the effect that the properties ascribed by the concepts introduced in the story—and so, plausibly, by their expressive counterparts in our language—are naturalistic in character. Insofar as the reconstructive analysis pairs in this way with the more familiar reductive model, it can be seen as seeking to realize the same goal by somewhat different means.

Does this account of the reconstructive approach simplify things in a parallel manner to that in which our account of the reductive does so? Yes, insofar as there are plausible variations on the genealogy that would generate somewhat different concepts and somewhat different candidates for expressive equivalence with our own ethical concepts. But any genealogy will tend in the nature of the enterprise to support a more or less determinate set of candidates for the role of expressive counterparts to our familiar concepts. It will correspond to a reductive account that has already selected the intuitive conditions that should be privileged in analysis of the concepts. The narrative developed here is determinate in precisely this respect. It is possible that plausible variations on the narrative might direct us to somewhat different expressive counterparts. But the possibility will not be explored here.[10]

Because of how they correspond, the reductive and reconstructive methodologies share a certain vulnerability. The vulnerability in the reductive case is that for all that has been argued, a property that is shown to be capable of being grounded naturalistically may not actually be **(p.24)** grounded in that way. The vulnerability in the reconstructive is that a property that is shown to be capable of having been accessed and conceptualized on a naturalistic basis may not actually be accessed and conceptualized in that manner.

How serious is this vulnerability in the analysis of the target concept and property, whether the analysis be reductive or reconstructive? Not very serious, I think. If we have a naturalistic reduction or reconstruction to hand, it would

offend against parsimony to hold that nonetheless the concept refers us to a non-naturalistic property and demands a non-naturalistic analysis. Why multiply entities by positing that property, when by hypothesis the concept makes perfectly good sense in naturalistic terms?

Given the choice between reduction and reconstruction, what reason is there to seek a reconstructive rather than a reductive account of ethical properties, or indeed of any naturalistically problematic properties? There is a consideration of methodological accessibility that argues in favor of the reconstructive approach with problematic properties in general. And, more to the point of the current project, there is a consideration of explanatory value that argues in favor of the approach with problematic concepts and properties of the kind involved in ethics.

The advantages of reconstruction: easing the methodological burden

The consideration of methodological accessibility is that reconstructive analysis is likely in most cases to be easier to achieve than reductive analysis. There are two respects in which this is so, and they are associated in turn with the two claims that each must make.

The first of these claims concerns the shape of the target concept: in the reductive case, the claim is that the concept applies in such and such conditions and does not apply in others; in the reconstructive, the claim is that a concept employed within the story of emergence is expressively equivalent, or more or less equivalent, to the target concept. The advantage of the reconstructive approach on this front is that the exercise it requires is second nature to all of us, adept as we must be in determining **(p.25)** whether we are thinking in broadly the same terms as others about any topic: whether we are using those terms under the same prompts and to the same purposes. By contrast, the exercise required by the reductive analysis is notoriously challenging and divisive: as noted, it involves using the method of cases, in which we seek agreement in intuition about the range of possible scenarios where we would apply the concept and about the corresponding range where we would not.

The second claim that each approach, reductive or reconstructive, must make is that there is a naturalistic candidate for the property that the target concept predicates. The advantage of the reconstructive approach on this front is that it only has to show that the story invoked is naturalistic in character, since that implies that there must be some naturalistic property or configuration, maybe this, maybe that, which is ascribed by the target concept. What the reductive approach has to do, however, is to identify the sort of naturalistic property that can serve as a plausible referent for the concept.[11]

The advantages of reconstruction: elucidating practice-dependent concepts

But apart from its greater methodological accessibility, the reconstructive approach also appeals on the ground that it has a particular explanatory advantage in making sense of certain sorts of concepts and properties. These are properties that become salient, and concepts that become accessible, only from within various practices, and include properties and concepts of the kind associated with ethics.

We noted earlier that it is not enough for the appearance of ethics that people generally act in conformity to the demands of moral concepts; it **(p.26)** is necessary that they have the concepts that support such conformity and are able to deploy them in typical ethical practices. These practices include the evaluation of various types of behavior as well as exercises like granting claims to others or making claims upon them, exhorting ourselves or others to live up to certain standards, and holding ourselves and one another to those standards.

The concepts and practices involved in ethics are interdependent. The practices presuppose the availability of the concepts, since it would make no sense to imagine people assessing types of behavior, assigning claims to one another, or enjoining one another to act in certain ways, if they did not have access to corresponding concepts. But equally the concepts would scarcely have a role to play in human life unless people were positioned to engage in practices of that kind. The practices are concept-dependent, as we might say, the concepts practice-dependent.

The practice-dependence of a certain class of concepts suggests that in order to gain competence in the use of any concept—in order to be sensitized to the pattern that the concept articulates—you would have to know how to engage in the relevant practices. In order to be sensitized to the pattern articulated among chess players in the concept of check-mating or castling, and in order to have a proper understanding of the property involved, you would have to know how to play chess; you might not have to be an expert, but you would at least have to know the rules: you would have to be on the inside, as it were. The lesson here, in parallel, is that in order to gain access to ethical concepts you have to know how to engage in standard ethical practices. It is only those with inside knowledge of the rules of chess who can properly master the concepts of check-mating and castling. And it is only those with inside knowledge of the practices of ethics who can properly master moral concepts.

This creates a difficulty for the reductive enterprise. A reductive analysis of check-mating or castling would identify the property ascribed by the concept from the outside, as a property that plays a certain role in the course of a game of chess. And a reductive analysis of ethical concepts would identify the properties they ascribe from the outside, as properties that serve in certain roles within ethical practice. But such **(p.27)** an outside analysis would communicate

only a parasitic understanding of a concept. It would identify the property ascribed by the concept in the way a color-blind person might identify the property of redness or blueness: that is, by the role it plays in the discriminations of those on the inside; in this case, those who can see color. It would not communicate anything like an insider sense of the property at issue.[12]

This limitation on reductive analysis does not affect reconstructive analysis in the same way. A reconstructive analysis of ethics must build on a narrative that explains why ethical concepts and practices should have co-evolved in a series of stages. The reconstructive analysis of money that we mentioned in the introduction does precisely this. It describes how monetary transactions and concepts might have co-evolved in stages, with exchange practices developing only in the presence of suitable concepts and the concepts emerging only in the presence of suitable practices. What we may hope for in the case of ethics or morality is that such a story of co-evolution may be available here as well, in particular a story of co-evolution that invokes only naturalistically intelligible adjustments.

In telling a story of co-evolution, the reconstructive analysis will have to appeal for its confirmation to our capacity to simulate how things would present in the wake of evolving developments of practice, and how insiders to those developments would find certain patterns salient and could evolve terms or concepts to articulate them. This is what happens with the genealogy of money, for example, when we are invited to imagine how gold or cigarettes would begin to present, as they are invoked by people in a barter society to serve as a generalized medium of exchange, and get to be used then as a metric for pricing things and as a means of storing wealth. The role of simulation in reconstructive analysis means that with ethics as with money, it ought to give us a simulated sense of the patterns articulated in moral concepts; and this, **(p.28)** without forsaking the naturalistic ambitions that reconstruction shares with reduction.

This particular advantage in reconstructive analysis should become clearer in the course of this book. It will be addressed again in the final chapter, when we look at the role of various practices in revealing moral properties. And it should be salient at two points in the narrative itself. One is in chapter 5, when we argue that as the protagonists avow desires rather than just reporting them, the patterns associated with various concepts of desirability will assume an inescapable salience for them. And another is in chapter 6, when we defend the view that in virtue of relying on one another’s pledges of fidelity to certain standards, the protagonists will have to be prepared to hold one another to account—effectively, to hold one another responsible—for how well they do by those standards.

A social reconstruction

A reconstructive analysis might be developed on a wholly individualistic basis with certain concepts and properties, even with concepts and properties of an interdependent character. In that case it would consist in a story of how a single individual might adjust to certain circumstances in a distinctive manner, how that pattern of adjustment might provide a base for thinking in a certain way, how that mode of thinking might create new possibilities of adjustment, how this in turn might lead to a more nuanced form of conceptualization, and so on.

Ethical concepts and properties would scarcely be amenable to such an individualistic account, however. The concept of desirability serves individuals in making sense of their choices to others and in responding suitably to certain complaints. And the concept of responsibility serves people in a parallel way to commend or condemn the choices of others or to express corresponding responses to their own decisions. The social role played by the concepts suggests that if we are to tell a plausible story about how people might come to master such concepts, then we must give an account of how they might develop the concepts in **(p.29)** tandem with one another: that is, as the members of an interactive community.

This means that our reconstructive analysis has to aim at providing a narrative about how a community of individuals who do not initially employ ethical concepts might evolve communal practices to a point where such concepts would become available to them and corresponding properties become accessible. The narrative would have to start with a possible, naturalistically credible form of human society—ground zero—where the members do not yet have access to any ethical concepts, in particular any concepts in the family of desirability or responsibility. And it then would have to show how naturalistically credible adjustments would give rise to certain patterns of interaction and lead protagonists to develop concepts that can be seen as more or less equivalent to the ethical concepts we routinely invoke.

In order to emphasize that ground zero in the narrative envisaged is not our actual social world, I give it the name of Erewhon. This name, borrowed from a nineteenth-century novel, is an anagram of “nowhere” and may serve as a reminder of the unhistorical nature of the community with which the narrative has to start. The task is to identify naturalistically intelligible and plausible developments among the residents of Erewhon that would lead them, without their necessarily being aware of the process, toward the introduction and mobilization of concepts akin to our concepts of desirability and responsibility.

The narrative about how ethics could emerge in Erewhon will be developed in chapters 2 to 6. There are two goals in this narrative, corresponding to the two claims associated with reconstructive analysis. The first is to tell a story in which the members of Erewhon come to think in broadly ethical terms, using those

concepts under broadly the same prompts, and to broadly the same purposes, as we use concepts like those of moral desirability and responsibility. And the second is to give such a manifestly naturalistic explanation of the concepts and practices that appear in Erewhon—and hence of our corresponding concepts and practices—that there is little option but to assume that they are naturalistically intelligible.

(p.30) Factual history, counterfactual genealogy

Given that the narrative provided offers a history of ethics in Erewhon, a natural question bears on the relationship between the story told and actual histories of ethics. Theories that purport to tell us about the actual history of ethics sometimes offer accounts of broadly ethical patterns of desire and behavior without paying much attention to the emergence of ethical concepts.[13] Those theories are not of much concern from the point of view of the current project, however engaging and even compelling they may be in themselves.[14] But how does our reconstruction relate to histories of ethics in which behavioral and conceptual patterns both make an appearance?

In a prominent example of such a history, Michael Tomasello (2016, 154) offers an "imaginative" account of how ethical ways of acting and thinking emerged, building that story on two sources of data: first, evidence about the stages in which our forebears evolved, especially from about four hundred thousand years ago; and second, evidence about the predispositions present in young children, but not in other great apes, for which our ancestors were likely to have been selected in that period. Tomasello's history is imaginative or speculative because, as he says, there is "little in the way of artifacts or other paleoanthropological data to help" (154). So how is the account developed here likely to connect with his enterprise?

Both projects aim at explaining ethics rather than explaining it away in a debunking fashion. But there is a sharp contrast between them. Tomasello explores the emergence of ethics in the actual conditions in **(p.31)** which early humans presumptively operated—his main focus is on the period between four hundred thousand and one hundred and fifty thousand years ago—whereas ours looks at the emergence of ethics in the conjectural conditions that characterize Erewhon. Given the assumption that actual conditions were as he describes, Tomasello offers an account of how ethics might possibly have emerged there. Given the assumption that ethics evolved socially, we in contrast offer an account of conditions that would have made their emergence more or less inevitable.[15]

This contrast between the projects derives from a difference in their goals. Tomasello's aim is to excavate the origins of ethics by looking at factors that might possibly have given rise to ethical ways of thinking and acting in the actual history of our species. The aim here is to explore the nature of ethics by

looking at factors that would almost certainly have given rise to ethical ways of thinking and acting in the conjectural conditions of Erewhon. Where the first aims at an actual history, the second aims at a counterfactual genealogy.[16]

If this account of the relationship between the two projects is correct, then there is room for fruitful exchange between them. In the project pursued in this book, for example, there are assumptions about the mutual reliance essential to the inhabitants of Erewhon that are strongly supported by the argument, prominent in Tomasello, that the early humans among whom morality emerged lived under conditions where, unlike other primates, they were required to forage together or die **(p.32)** alone. And if the project pursued here is sound, then it may be that Tomasello should make more use of the role of language in order to explain how our ancestors came to conceptualize and govern their relationships in ethical terms (Pettit 2018b). But the possibility of such cross-fertilization is not our topic. Putting aside the actual history of ethics, the task is to make sense of the project of reconstructive analysis.

1.2 Conditions on a successful reconstruction

If a reconstructive analysis is to serve the role envisaged for it here, then there are some conditions, as already indicated, that it has to meet. It must satisfy the input condition of starting from a possible society that is naturalistically intelligible and, despite lacking ethical and related concepts, is fairly similar to ours. And it must satisfy the process condition of explaining in broadly naturalistic terms why certain evolutionary developments would materialize in that world. We will examine how far these conditions constrain the genealogy to be presented and look then at the way in which it contrasts with parallel approaches that might be taken in the reconstruction of ethics.

Introducing the input condition

The main input requirement on our narrative has to be that the members of Erewhon initially lack access to ethical concepts like those of desirability and responsibility. But it will be useful to stipulate that they also lack access to other prescriptive concepts, such as those associated with desirability under law or etiquette or epistemology, or with desirability in prudence or in patriotism. This stipulation makes the task more challenging and interesting, since allowing participants access to other prescriptive concepts might make it suspiciously easy to explain their developing a concept of the morally desirable.

In order to have a good chance of being feasible, however, the reconstruction must take Erewhonians to be as close as possible to human beings like you and me, despite not yet having prescriptive concepts at **(p.33)** their disposal. Thus, the project would almost certainly be infeasible, if it started with creatures, like some of our evolutionary ancestors, who did not yet form societies, or that did not depend on establishing social relations with one another for achieving

individual success, or that did not have the cognitive capacities required for natural language.

Reflecting this constraint, the reconstruction provided here assumes that Erewhonians are very like us on a variety of fronts. First, of course, they have beliefs and desires and act routinely for the satisfaction of their desires according to their beliefs. Second, even if they are moderately altruistic, they primarily desire the promotion of their own welfare and that of their kin. Third, they are able to rely on others, and able to get others to rely on them, as an essential means to promoting that end.[17] Fourth, they have the capacity in pursuing mutual reliance, first, to exercise joint attention, consciously focusing on data they take to be available to all, and second, to act jointly with one another in pursuit of shared goals (Tomasello 2014).[18] And fifth, they are able to build on those capacities and use words in the communicative fashion of natural human language (Scott-Phillips 2015).

While the inhabitants of Erewhon have beliefs and desires, however, these attitudes do not involve any prescriptive properties in their contents: for example, they do not include beliefs about what is desirable or, equivalently, about what there is reason to desire or what they ought to desire. And while they can use natural language to express their attitudes, they use it only for the limited purposes of giving one another reports on how things are in their environment. They communicate about whether the blackberries have ripened on the hill, about what the weather is like farther north, about how the prospects are looking for the big-game hunt. And that is all. Language serves them only as **(p.34)** a means whereby they can trade information, to their mutual benefit, about the world they occupy.

It is unlikely that there ever was a time or place in the trajectory of human development, of course, when members of our species used language solely for making reports on their shared world. But, anticipating later discussion, the claim to be developed in the reconstruction is, first, that if Erewhonians used words to communicate about the world in this manner, they would also use them to communicate about their own attitudes; second, that in communicating about their attitudes they would be more or less bound to rely on speech acts of avowal and pledging, co-avowal and co-pledging; and, third, that with those practices in place, they would naturally use certain terms under such prompts and for such purposes that the terms count as expressively equivalent to our concepts of moral desirability and responsibility.

The world that the story posits is not only meant to be akin in these ways to ours; it is also taken, as a presumptive aspect of that kinship, to be intelligible in naturalistic terms. That means at a general level that it should be a world intelligible on the basis of the properties recognized in natural science or grounded in properties recognized in natural science. And it means, in

particular, that it should be intelligible without postulating properties of the kind associated with prescriptive properties.

Interrogating the input condition

We may assume for present purposes that the postulation of intentional and linguistic competence in our protagonists does not offend at a general level against the assumption of naturalism: that whatever form that competence involves, it is grounded in properties familiar from natural science; it does not presuppose a Cartesian, immaterial mind, or anything of the kind. But does the assumption that members of Erewhon have intentional competence, and competence in using natural language to express their intentional attitudes, mean that they are already living under a prescriptive regime? And does it offend in that particular way against the claim to be able to characterize Erewhon in presumptively naturalistic terms?

(p.35) Intentional competence means competence in the formation of attitudes: that is, a capacity to form attitudes under appropriate conditions—say, to form a belief that p in the presence of evidence that p—and a capacity to meet constraints of consistency and the like in doing so. And linguistic competence means a capacity to meet corresponding constraints on the use of language, in the event of wanting to express those attitudes. Such intentional-linguistic competence may assume a low-grade or a high-grade form: it may involve rational processing or reasoned argument. But, whatever shape it takes, it does not require our protagonists to operate under a prescriptive regime; it does not require them to have access to concepts of desirability, moral or otherwise.

Low-grade competence would require them to conform to suitable constraints or regularities in the formation of intentional attitudes but to do so without any degree of intentional control: to do so in an unconscious, non-intentional form of rational processing. High-grade competence would require them in addition to recognize, now on one occasion, now on another, that this or that response is required under relevant constraints; to have a disposition on suitable occasions to respond as required; and to be able to take measures, however indirect, to promote the likelihood of their responding appropriately. It would require reasoned argument rather than mere rational processing.

In order to display the first sort of competence, our protagonists would just have to be disposed to transition rationally from beliefs about the satisfaction of suitable conditions to the formation of other beliefs and to be disposed, depending on what they desire, to give expression to such beliefs. And they could do this just as unconsciously or sub-personally, and just as automatically, as the well-designed robot. Forming the belief that p and that if p, q, for example, they would automatically conform to modus ponens by forming the belief that q and, if they wished, would express that belief in words. They would do this, in particular, without having any desire to satisfy the requirements of

the rule and without having any intentional control over whether or not they do satisfy them.

In order to display the second, high-grade sort of competence, our protagonists would have to be able to reason their way to the conclusion **(p.36)** that q, not just transition rationally into holding that belief. In reasoning to the conclusion, they would pay attention to the state of affairs that holds according to their premise beliefs—pay attention to the fact, as they see it, that p and that if p, q—and then form the belief that q in conscious response to that presumptive fact. In such an event, they might think "That implies that q," or simply "So, q." They would not just have a belief in the premises that generates a belief in the conclusion; they would think of the truth of the premises in which they believe as forcing them—as it happens, under the rule of modus ponens—to acknowledge the truth of the conclusion and to form a belief also in it.[19]

This ratiocinative, high-grade competence involves taking relevant rules, if only in a case-by-case way, as normative guides (Brandom 1994; Wedgwood 2007). Those rules will include not just a rule like modus ponens but also rules that express the constraints of evidence, consistency, closure, and the like. The capacity for reasoned argument consists in the ability to take intentional steps—say, in exercises of attention and care—to conform to such rules: to practice rule-following.[20]

Low-grade rational processing would certainly not require our protagonists to operate under a prescriptive regime, deploying concepts like those of desirability. But would the presence of reasoned argument mean that they would have had access to such concepts and would have been positioned to prescribe the satisfaction of relevant constraints to one another and to themselves?

No, it would not. In reasoning in the manner of modus ponens, the protagonists would have to treat the presumptive fact that if p, q and that p as a reason for believing that q; they would do this in registering and responding to the pressure to believe that q—that is, in forming **(p.37)** the thought expressed in, "So, q." But they might treat that fact as a reason for believing that q without developing the concept of a reason and without having related ideas about what is evidentially compelling or epistemologically desirable. Thus, it is possible to reason from certain premises to a conclusion without having access to the idea that the premises provide a reason for holding by the conclusion. For current purposes, it is enough to vindicate this possibility only for the theoretical reasoning that prompts the formation of belief; we shall turn later in the chapter to the parallel possibility for the practical reasoning that eventuates in the formation of desire or intention.

The claim that this form of reasoned argument would not put our protagonists within reach of a notion like that of epistemological desirability is supported by

later discussion. We shall see in chapter 5 that a concept of the evidentially credible would emerge on the basis of the same sorts of practices supporting the emergence of other concepts of desirability, including moral desirability. The salience that those practices would give to the prescriptive notion of credibility suggests that nothing like it would be salient, or even perhaps available, in their absence.[21]

Introducing the process condition

As the possibility of success requires the satisfaction of various constraints on the input side of the reconstruction, so it also requires the satisfaction of certain constraints by the process that the story describes. If the reconstruction is to be successful, this requires, first, that it does not introduce any non-naturalistic factors into the story: it must not postulate a naturalistic miracle; second, that it does not rely on any **(p.38)** fortuitous occurrences: it must not postulate lucky flukes; and third, that it is the most economical story available: it must not be an idle wheel. Lacking any of these features, it would not direct us to a naturalistically plausible reconstruction.

In order to be naturalistically intelligible, to take up the first constraint, the process cannot presuppose elements or forces that are ungrounded in natural science. And equally, of course, it cannot presuppose that the protagonists already have access to prescriptive properties. This constraint is as clearly needed, and as relatively unproblematic, for the process condition as it was for the input condition. But why require in addition that the process be non-fortuitous and economical?

If the narrative invoked in explanation of the emergence of ethical concepts in Erewhon relied on the occurrence of something antecedently unlikely and serendipitous—if it relied on a lucky accident—then it would have the character of a just-so story. All that it would establish is that, as naturalistic factors lead by a lucky fluke to the emergence of recognizably ethical concepts in Erewhon, so the concepts that respond to similar prompts and serve similar purposes in our ways of thinking may also have emerged by a lucky fluke under the impact of naturalistic factors. This result might have a certain interest in establishing the possibility that our ethical concepts predicate naturalistic properties, but it could hardly establish that such naturalistic properties are prominent candidates for the properties that those concepts actually ascribe.

This explains why the narrative should be non-fortuitous as well as naturalistic. But, turning to the third constraint, why assume that it must also be economical? Why assume that it must be the most parsimonious narrative available?

The narrative to be presented here makes language essential for morality; it assumes that before developing prescriptive concepts, and accessing prescriptive properties, the protagonists already communicate linguistically. This

account would not be parsimonious, intuitively, if there were another equally plausible narrative available under which the protagonists became aware of suitably prescriptive properties without having access to language and without first having terms in which to predicate them.

(p.39) Assuming that this alternative story offered different candidates for the role of the relevant prescriptive properties, both in the target language and in ours, there would be good reason to prefer it.[22] That account would link prescriptive predicates with properties that would have been salient to the protagonists, even if they had never developed language, and had never introduced terms to predicate the properties. And those properties, presumably, would be salient to us on a firmer basis than the rival set of properties and would be more prominent candidates for the referents of our own terms; they would not depend for their accessibility on the effects of induction in language.

How do the three constraints just rehearsed shape the narrative to be offered here? The first constraint will be satisfied to the extent that none of the developments introduced in the story presupposes any non-naturalistic capacities in the protagonists or access to any prescriptive properties. And the third constraint will be satisfied, at least presumptively, to the extent that no more parsimonious but still naturalistic alternative is in the offing; the narrative is presented under the implied tag line: 'if not this, what?'.[23] But what of the second anti-fortuity constraint, against the introduction of lucky flukes?

Unlike the other two, this constraint has a direct impact on how the narrative is constructed. It requires us to offer a characterization of the psychology and circumstances of the inhabitants of Erewhon that enables us to tell a story as to why they would develop ethical concepts that is not unrealistically rigged in favor of the development. The assumptions it makes should be realistic or, if not fully realistic, they should make it harder rather than easier to explain the emergence of morality; they should posit a worse-case rather than a better-case **(p.40)** scenario for its emergence. The assumptions governed by the constraint bear, first, on the psychological profile of protagonists in the story and, second, on their anthropological circumstances.

Interrogating the process condition: the psychology postulated

The psychological assumptions made in the narrative presented here depict the inhabitants of Erewhon as rational actors who, notwithstanding a degree of altruism, are disposed to try to avail themselves of any salient opportunities—including opportunities involving mutual reliance—for bettering their lot. The reason for relying on this rational, relatively self-regarding profile is that it offers a firm basis on which to predict the actions and adjustments of the protagonists fairly reliably.

But it may not seem realistic to postulate that the inhabitants of Erewhon are rational or that they are self-regarding. Or at least this may not seem realistic in view of the assumption that since they are psychologically like you and me, they must be susceptible to various failures of rationality, on the one side, and that they must be capable on the other of various forms of altruistic concern.

Taking up the rationality issue first, it is now common to observe that human beings like you and me fall short of full rationality in a variety of ways. Thus, for example, we rely on a battery of biases and heuristics that generally made evolutionary sense, at least in ancestral environments. And these lead to irrational responses in a range of familiar circumstances (Gilovich, Griffin and Kahneman 2002).

Despite this consideration, however, there are two features of the rationality on which the narrative relies that make it a realistic postulate. The rationality postulated is limited in degree, appearing mainly in the disposition of agents to prove reciprocally reliable to those on whom they themselves rely. And it is a sort of rationality that agents are required to display in recurrent situations, so that they will be in a position to confirm its benefits, time and time again.

(p.41) The second reason to doubt the psychological profile ascribed to our protagonists is that human beings are not as self-regarding, by many accounts, as the narrative supposes. It is plausible, as Michael Tomasello (2016) argues, that as a result of their self-regarding needs, Mother Nature selected our forebears for the presence of proximate psychological mechanisms triggering mutual cooperation and reliance quite spontaneously; psychological studies of children show that such dispositions are in evidence even before the age of three. By this account, our ancestors would still have been moved, as we continue to be moved, by self-interested concerns for them and theirs. But they would plausibly have developed a natural disposition to be cooperative, making and living up to joint commitments of various kinds. And, presumably, they would have evolved a disposition to share culturally accumulating customs and skills, passing these on to their young (Sterelny 2012).

Assuming that this is right, it is clear that the self-regarding aspect of the psychological profile ascribed to the inhabitants of Erewhon, unlike its rational aspect, is not fully realistic. But this need not be a problem. For the ascription of a primarily self-regarding profile to our protagonists makes it harder rather than easier to explain why they should develop an ethical mindset. This aspect of their profile means that the world tracked in our narrative is a worse-case rather than a better-case scenario. And so, if a recognizable ethics could have emerged there, it would surely have emerged in the presence of a lesser degree of self-regard. What our nature would have generated in the dry wood of Erewhon, it is all the more likely to have generated in the green wood of our actual history.

The narrative developed later suggests that practices of avowal and pledging will put members of Erewhon within reach of ethics by forcing them to single out properties that are robustly attractive, enabling them to sustain the desires and intentions they avow and pledge. One effect of ascribing a rational, self-regarding psychology to those members—one effect of not presupposing natural, cooperative predispositions—is that it may put limits on the sorts of properties they are individually likely to treat as attractors or desiderata of this kind. If Tomasello is right, for example, they are less likely than our forebears to take a property like **(p.42)** that of being helpful or cooperative to serve in itself as an individually attractive desideratum.[24]

But this difference does not put the current project in danger. The limitation on the psychology of members of Erewhon may affect the character of the ethics that evolves among them. But the fact that ethics would evolve among them may still throw light on its role and nature. Certainly, this will be so to the extent that we can recognize the moral terms and concepts they come to employ as expressively equivalent to counterparts in our own moral vocabulary and thought.

Interrogating the process condition: the circumstances postulated

These considerations should help to vindicate the psychological assumptions built into the narrative. But the requirement that the narrative should not rely on lucky flukes puts a constraint, not just on the psychology ascribed to members of Erewhon, but also on the circumstances postulated. As in the psychological case, the constraint requires that those circumstances should be realistic or, if not fully realistic, that they should not rig things in favor of the development of ethical concepts and practices.

In the story told here, Erewhon is an isolated society and a society in which members enjoy relatively equal power. The assumptions of isolation and equality facilitate the narrative in the same way as the assumption of a broadly rational, self-regarding mentality. There are many shapes that external relations might assume, only one that answers to isolation. And there are many forms in which inequality might appear, only one where there is equality. Assuming isolation and equality makes the project undertaken more readily feasible.

(p.43) But do the assumptions of isolation and equality rig the books unrealistically in favor of the appearance of ethics? Isolation is certainly not a problem. It is plausible that over long periods of our history, many communities of human beings lived in isolation from others, even in ignorance of the existence of other societies. And in any case the assumption of isolation does not make it easier in any obvious way to explain the emergence of ethics. It may even make it harder to do so, since contact with other societies might

foreground the role of ethical standards and highlight the attraction of establishing them across communities.[25]

The assumption that the members of Erewhon enjoy relative equality of power raises a sharper challenge for the process condition than does the assumption that it is an isolated society. For this assumption does seem to make it easier to provide an explanation of ethics, or at least an explanation that does not debunk that which it explains.[26] But there are two things to say in response to the challenge.

The first is that the assumption of relative equality is not unrealistic to the point of positing a lucky fluke. For a high degree of equality appears to have prevailed in most of human history: specifically, in the one hundred thousand years or so prior to the agricultural revolution, which occurred less than ten thousand years ago (Boehm 1999; Boix and Rosenbluth 2014). But even if this claim is put in doubt, there is a second, more important point to make in defense of the assumption.

(p.44) While the narrative developed here supposes a society in which all are equals across divisions like those of gender or class or ethnicity, it can serve the purpose for which we employ it under variations on that supposition that might be considered more realistic, if less appealing. I am thinking of the sort of variation in which women are subject to men, the weaker to the stronger, or those in a special enslaved category to a master class. If there is such a division, then the sub-society of the privileged is still likely to display a high degree of equality within it. And the narrative developed here for the inclusive society of Erewhon would be likely to work for such an elite, showing how they would develop ethical concepts—broadly, concepts expressively equivalent to ours—governing their treatment of one another.

How would the elite narrative serve the same purpose as the narrative for Erewhon? It would show that any relatively egalitarian society or sub-society would be likely, starting from something like our ground zero, to develop ethical concepts like ours and to evolve corresponding practices in dealing with one another. And that would support a naturalistic analysis and vindication of ethics, as envisaged here.

The narrative would do this, at any rate, to the extent that it did not essentially depend on the suppression of others outside the elite.[27] If it did essentially presuppose such suppression, then it would debunk morality as we think of it, representing it as a code appropriate only for a dominating elite. But if it did not essentially presuppose suppression, as presumably it need not, then it would present morality as a way of thinking and acting that can—and by our lights should—be extended to all. Or at least it would do this, on the assumption that

the concepts employed are expressively equivalent, or more or less equivalent, to our ethical concepts.[28]

(p.45) The characterization of the process condition in this discussion combines with the earlier characterization of the input condition to hold out the prospect of a persuasive narrative of development. If Erewhonians have defined, self-regarding purposes, live in isolation from other societies, and occupy positions of relatively equal power, then it is going to be feasible to think of charting the opportunities they are going to confront; to predict the responses they are likely to make to those opportunities; to see how those responses can aggregate to create further opportunities; to predict how they are likely to respond in turn to these; and thereby to trace a plausible trajectory of development. And that is precisely what a reconstructive analysis must seek to do, aiming to show that the trajectory charted is likely to culminate in the appearance of recognizably ethical concepts.

Three contrasting projects

By the account offered so far, the distinguishing mark of the ethical concepts of desirability and responsibility is that they are prescriptive. If the reconstructive analysis of ethical concepts is going to be successful, therefore, then it must explain the appearance of such concepts in the course of developments in Erewhon. And it must do so without presupposing the earlier presence of any ethical concepts—or, as we postulated, any prescriptive concepts at all.

This requirement implies that the reconstruction to be developed here should be distinguished from three more familiar forms of narrative. These are associated respectively with game theory, contract theory, and the sort of theory that presents ethics as prudence by another name.

The game-theoretical narrative is regularly invoked in computer simulations of the emergence of altruism, in psychological experiments that provide evidence of mutually beneficial adjustments among subjects, and in evolutionary explanations, natural or social, of the emergence of cooperation and altruism.[29] While these projects are often **(p.46)** of great interest in themselves, none of them aspires to play the role expected here of a reconstruction of ethics, since they do not posit, let alone explain, the emergence of ethical concepts. They focus entirely on explaining the emergence of attitudinal or behavioral patterns and do not attempt to explain how ethical concepts could have emerged and exercised influence in the ordering of people’s relationships. In the story told here, attitudinal and behavioral adjustments inevitably have an important place. But they figure in a narrative in which concepts co-evolve with the attitudes and behaviors postulated and interact with them in the maintenance of social life.

The need for a reconstruction to explain the appearance of ethical concepts de novo rules out, not only the game-theoretical alternative, but also the standard

sort of contractual story. The narrative is supposed to show that despite not having access to ethical concepts or practices to begin with, the inhabitants of Erewhon would be more or less inevitably pushed toward introducing them. But this means that they cannot establish a use for those concepts on the basis of anything like a contract that is represented as a contract to establish a morality. For in order to form such a contract with one another they would already have to possess the concept of what it is they are contracting into. And that means that the story would not explain the emergence of ethical concepts de novo.

The idea to be defended in the reconstruction, then, is not that the members of Erewhon would have motives to enter a social contract with one another for establishing moral standards, as political theory has often invoked people's motives to explain their entering into a political contract. To anticipate again, the proposal rather is that those individuals would be moved to avow and pledge their attitudes, rather than reporting them, and that with avowal and pledging established as shared activities, they would be moved in turn to develop properly ethical concepts. The narrative has to document an unplanned process of more or less inevitable emergence, not a history of contractual agreement. It has to be scripted in the spirit of David Hume (1978, 3.2.2), when he relied **(p.47)** on such a story to explain how we could have developed the concept and practice of promising and contract, on the one side, and ownership and property on the other.

The requirement of de novo explanation means, finally, that the reconstruction sought here cannot take the form of a story under which individuals each come to reason on prudential grounds—on grounds of their individual, long-term interest—that they should embrace ethical standards of desirability and responsibility: they should recognize, in a variation on a cliché, that morality is the best policy. Such a prudential story would presuppose the availability of ethical concepts; in order to argue that morality is the best policy, after all, you have to have the concept of morality. Hence this approach would fail in the same way as the contractual narrative.[30]

As characterized here, the contractual and prudential stories are stories as to how those who already have moral concepts available might mobilize conformity to moral requirements on a contractual or prudential basis. But might such stories not posit independently intelligible contractual or prudential adjustments —that is, adjustments that do not presuppose access to moral concepts—and argue that those concepts would appear in the wake of the adjustments? Yes, they might. But such alternative stories would be akin to stories of emergence of the kind envisaged here; they would not constitute real alternatives. The prudential story would differ at most in positing access to a prescriptive concept of self-interest; and the contractual story would diverge in positing a more demanding, collective form of agreement than that which is generally envisaged

in the current narrative. They would each offer an alternative narrative of emergence, albeit ones of a less parsimonious character.

The role of reasoned, practical argument

The contractual and prudential theories suggest that the parties involved reason their way to a conclusion in favor of ethics. Even if we **(p.48)** reject those theories, that alerts us to a question that we should address in conclusion. The story represents the protagonists as identifying and exploiting various opportunities for advancing their ends, as we have seen. And so, the question is whether it can countenance the exercise of instrumental reasoning in the process of working out the best means to those ends. Does it allow that in making presumptively rational responses to the opportunities they face, the members of Erewhon may reason their way to conclusions about what to do? Does it allow this, in particular, from the initial stages of the narrative?

We know from earlier discussion what it means to reason theoretically, in accordance with modus ponens, that since it is the case that p and that if p, q, it must be the case that q. But what does it mean to reason practically, in accordance with instrumental rationality, that since the end is E, and X-ing is a way—perhaps the only way—of getting E, X-ing is the thing to do?

In theoretical reasoning, our protagonists do not merely form a belief in the premises and then, as a result of an unconscious, sub-personal mechanism, form a belief in the conclusion; they think of the truth of the premises in which they believe as forcing them to acknowledge the truth of the conclusion, and they form a belief in the conclusion under that perceived pressure. The practical reasoning required in the instrumental case must also involve more than an unconscious form of rational processing. It is not going to be enough that the presence of a desire for E together with a belief in the proposition that X-ing is a way to realize E unconsciously and automatically generates a desire or intention to X.

What happens in the case of reasoned, practical argument must parallel the sort of thing that marks off reasoned argument in the theoretical case. The protagonists presumably have to pay attention to the state of affairs that holds when, for example, they desire E and believe that X is a suitable means to that end; they have to see this as pressuring them to form the desire to X; and they have to respond by forming that desire. They have to form the desire in such a way that they can be represented as thinking something like: “so let me X,” or “so X-ing is the thing to do.”

(p.49) Unlike the sketch of theoretical reasoning, however, this schematic account of practical reasoning communicates little or nothing about the state of affairs to which the protagonists are supposed to attend and respond in finding their way to action. What is the state of affairs that holds if they desire E and

believe that X is a suitable means of realizing it? That state of affairs cannot involve E's being desirable, since by hypothesis the protagonists do not yet have access to any property of desirability. So, what are we to say? Are we to hold that they exploit the opportunities for advancing their ends solely on the basis of an unconscious, non-intentional form of rational processing?

This would not be an outlandish position to hold, since rational processing is all that agency strictly requires; it is all that decision theory assumes, for example, in characterizing rationality in general (Pettit 1991b). But short of ascribing prescriptive concepts to them, we can still imagine that the members of Erewhon may often conduct a conscious, intentional form of deliberation in exploiting the opportunities they detect and, in that sense, may reason their way to the conclusions on which they act.

Suppose that although they lack prescriptive concepts, the members of Erewhon still identify certain attractors or desiderata; this assumption will be supported in the course of our narrative. These are properties that tend to make any scenario attractive to them: that it would be fun, for example, or a source of food, or a way of protecting their family. The availability of concepts for such attractor properties would enable the protagonists in our story to conduct a sort of deliberation that answers broadly to the schematic notion of reasoning illustrated in the theoretical case. It would enable them to pay attention to the state of affairs that consists in E being fun and, X-ing being a way to realize E, to respond to that situation with an attitude formed in a way that might be expressed in the words: "so let me X."

With these points in place, we can think of the members of Erewhon, even at the start of our narrative, as relying often on reasoned argument, and not just on rational processing, in deciding what to do. They may not have access to prescriptive concepts of desirability, strictly **(p.50)** speaking, but that does not rule out the possibility of practical reasoning of the simple kind illustrated and of the many variations it would presumably allow.

1.3 Reconstructive analysis illustrated

In order to make the idea of a reconstructive analysis more vivid and perhaps more acceptable, it may be useful in conclusion to show how it is used elsewhere. Perhaps the most familiar case of reconstructive analysis in roughly our sense is provided by the genealogy of money mentioned in the introduction. But there are also other prominent examples of the approach, including examples within philosophy itself.

The reconstructive analysis of money

The most familiar analogue to the project undertaken is the account of money that is standard in economics (Menger 1892). Where our project seeks a naturalistically intelligible story about the emergence of morality, this account looks for an individualistic story about the appearance of money. The account

begins from a conjectural society in which members only conduct barter exchanges—by many accounts, no such society ever existed (Graeber 2011)—and uses that starting point to develop a reconstructive analysis of money in independently intelligible terms.

In the barter society imagined, which is an isolated, relatively egalitarian community like Erewhon, people are interested in exchanging various commodities or services but cannot easily find suitable partners. You want the dog that I can provide, but I do not need the service that you would give me in recompense. I want something that a third person can furnish, but that individual does not want my dog or anything else I can currently offer. People in such a society might improve things by writing IOUs in a suitable domain—for example, in the provision of puppies—but this would have similar, if looser limitations. So, what might relieve them of the problem they face?

(p.51) The standard story is that at a certain point it is very likely that some commodity like gold or cigarettes or cattle would assume a special status, being recognized as a commodity that a great number of people apparently want. At that point, it would be in the interest of each to gain access to that special good or to IOUs issued by individuals or groups who could provide it. People could be sure of finding suppliers for the things they wanted if and only if they had enough of that good, or at least of reliable IOUs in that good. Even if the suppliers did not want it themselves, they would want it for the fact that they could trade it with customers who wanted it, or with customers who themselves had customers who wanted it, and so on.

With these developments, that good and the corresponding IOUs would come to play a distinctive role in the imagined society. And as we reflect on the role they would play, we readily conclude that this is a role played in our actual society by money, so that to say that something is money—to ascribe that property to it—is just to say that it plays the required role. The role is displayed in the part played by the preferred good, or IOUs in that good, as a medium of exchange, as a metric for putting prices on things and as a means of storing wealth.

As we identify further likely or possible developments in our reconstruction, we can bring out other roles that money might play as well, whether as a matter of definition or as a contingent fact. The government might accept the favored commodity in payment of taxes, for example, giving it the status of legal tender; the issuers of IOUs might be legally recognized as banks; the supply of IOUs might be controlled by a body—in our terms, a central bank—that guards against over-supply and under-supply; and these IOUs might come to be backed solely by their trading value, not by the guarantee that owners can cash them in for a corresponding commodity like gold.

This narrative demystifies the practice and concept of money in independent terms—in the terms of economic theory—as our projected narrative would demystify the appearance of ethics in naturalistic terms. It explains how a practice and a concept would emerge in the envisaged scenario that respond to broadly the same prompts and serve broadly the same purposes as our practice and concept of money. And it **(p.52)** explains how this process of emergence would materialize in economically plausible circumstances via economically plausible steps, thereby making it likely that the property that is ascribed in the emerging concept, and presumptively the property ascribed in our counterpart concept, is grounded in economically intelligible properties. It is a bona fide property that belongs in common to coins, notes, checks, and all forms of credit.

A philosophical project

As reconstruction can serve in the analysis of money, so it can serve in analyses that are more characteristic of philosophy. There are a number of philosophical exercises in which we can see the method of reconstructive or genealogical analysis at work.

Think of the example mentioned in the introduction: Herbert Hart's (1961) account of how a spontaneous, social regime of primary rules would have required secondary rules for its maintenance and how this could have generated a recognizably legal form of practice and conceptualization. Think of Wilfrid Sellars's (1997) myth of Jones, according to which we could have developed concepts of mental experience and attitude, and begun to practice folk psychology, by seeking a theoretical explanation for our dispositions to make certain utterances and to take corresponding actions. Think of David Lewis's (1969) demonstration that as self-interested rational agents we could have coordinated with one another in familiar predicaments, and given rise to regularities of the kind that answer to the concept of a convention. Think of Saul Kripke's (1980) story about how we could have introduced names as causally linked tags for the things we name, without sharing any single view of the descriptive character of the things named (Jackson 2004). Think of Edward Craig's (1990) claim that we could have developed the concept of knowledge, and the practice of justifying claims to knowledge, out of an interest in determining who should count as good informants by criteria available to everyone in the community (Fricker 2010). Or think of Bernard Williams's (2002) explanation of how a community of mutual informants could have evolved norms of **(p.53)** truth and truthfulness without relying on any prior sense of a truth-telling obligation.

All of these projects, like the narrative about money, identify certain activities or practices that would have pushed participants in a certain direction, would have naturally prompted a cascade of adjustments, and would consequently have provided an occasion for the introduction of certain terms or concepts. The activities that figure in our examples are those of overcoming difficulties in the

application of spontaneous, social norms; seeking to explain one another's utterances; coordinating with one another in certain predicaments; finding tags by which to track particular items in conversation; identifying speakers who count as reliable informants; and testing for credibility in exchanges of information.

As in the narrative about money, people would have naturally made certain adjustments in pursuing these activities, would have been exposed as a consequence to corresponding patterns, and would have developed terms and concepts to mark their presence. In particular, they would have developed concepts that function like our familiar concepts of positive laws, mental states, social conventions, proper names, knowledge claims, and truth-related norms. And plausibly, the properties that those concepts are used to ascribe are good candidates for the properties our own concepts predicate.

Being developed on these lines, the projects serve to demystify the concepts they address. They propose respectively that the concepts developed within the narratives they tell—concepts that are taken to ascribe bona fide properties—correspond to our familiar concepts of positive laws, mental states, social conventions, proper names, titles to knowledge, and norms of truth and truthfulness: they respond to similar prompts and serve similar purposes. And they explain the emergence of those concepts in suitably unproblematic terms, thereby supporting the claim that the properties they ascribe—like the properties ascribed in our counterpart concepts—are grounded in unproblematic bases.

What counts as unproblematic varies from case to case, and so the reconstructions also vary in the claims supported. They hold respectively that there is nothing normatively mysterious about how we can appeal to positive laws in assessing and regulating behavior; nothing **(p.54)** epistemically mysterious about how we can ascribe mental states to ourselves and one another; nothing individualistically unintelligible about our identifying and conforming to conventions; nothing surprising about our using proper names in common for items that we each are liable to describe in different, even conflicting terms; nothing about states of knowledge that makes them more puzzling than other states of mind; and nothing about our attachment to truth and truthfulness that requires an independent sense of the obligatory.

What these reconstructive stories seek to achieve in their respective domains, the story sketched here aims at achieving in the domain of ethics or morality. The idea is to demystify our ethical concepts naturalistically by explaining how corresponding concepts might have emerged in a naturalistically unproblematic way; and to provide a reason to think, therefore, that the properties ascribed by

our concepts, like the properties ascribed by their counterparts in the story of emergence, are grounded in unproblematic, naturalistic properties.

Reconstructive analysis and creature-construction

Not only do we find practitioners of reconstructive analysis among contemporary philosophers, however; we also find defenders of an explicit methodology of philosophical analysis that is close in character to the reconstructive approach. This is the methodology that H.P. Grice (1975b) described as creature-construction. Foreshadowed by Jonathan Bennett (1964) in his study of rationality, it is often invoked in contemporary work. Michael Bratman (2014) applies it in analysis of shared agency, for example, and Peter Railton (2014) in analysis of belief.

To take an example close to the project here, Bratman uses the methodology to make sense of what it is that we human beings ascribe when we speak and think of ourselves—or indeed of others—as acting on a joint intention for the achievement of a shared goal. The goal may be that of dancing a tango together, taking a walk with some friends or cooperating with others to rescue a drowning child. Joint action of this kind may be thought to be problematic in individualistic terms, requiring a plural bearer for the joint intention. But Bratman explains **(p. 55)** in painstaking detail how individuals might actually come in purely individualistic steps to act on a joint intention and display joint action, with each person identifying the particular part required of them and each playing that part under a shared assumption that others will also play theirs.

This explanation can be taken to show that what we ascribe in postulating a joint action by a number of people may be just the same sort of property that might be realized in the case described individualistically. And it shows this without suggesting that in actual life people only cooperate with one another on the basis of the very exacting thought processes attributed to the participants in the individualistic story. When you and I tango together, for example, we may be incapable of identifying what we each do except in terms that presuppose the joint action; what we each do may be as opaque to us as what our right and left hands do in tying our shoelaces. But the explanation is nevertheless illuminating. For what explicit intentions do in generating the joint action in Bratman’s model are presumably done in actual life—certainly in an example like that of tangoing—by more or less unconscious adjustments that take their place (Pettit 2017a).

The reconstructive method followed in this book differs in two ways from that of creature-construction. First of all, it looks back at a possible process of construction in the past rather than describing a possible process of construction that agents might follow in the present. And, second, the process that it looks back on is not an intentional process like the process imagined in creature-construction; rather it involves a series of adjustments among the relevant agents—in our case, a whole community—that accumulate to produce

an unplanned and unforeseen effect. But these differences are not of deep significance. The methods of creature-construction and reconstructive analysis are alike in exploring a possible generative process in order to make sense of the nature of an actual phenomenon.

The goal of this book, then, is to pursue a familiar style of philosophical investigation, albeit in an area where it has not been much employed. In the following five, progressively longer chapters, we will try to tell a narrative about our imagined Erewhon that, despite being **(p.56)** naturalistic in its starting point and in the successive steps it charts, generates an ethical community like that which you and I inhabit. This is a community, as we shall argue, where people think in terms of desirability and hold one another responsible for living up to desirable standards. When that narrative is complete, we can return in the final chapter to teasing out the philosophical lessons that it teaches about morality.

Notes:

(1.) The properties that are actually realized by one or another configuration of fundamental properties include some that cannot possibly be realized on a different basis, as in the examples given, and some that in principle can (Jackson 1998). For most naturalists, examples of the latter would include a property like holding a belief that *p*, which is actually realized on the basis of a subject's neural or electronic configuration, but might conceivably be realized in a Cartesian world where mental states are of a totally different kind.

(2.) For an excellent discussion of the challenge, see Enoch (2011, 100–109). According to a currently popular version of non-naturalism, there is no need to find a naturalistic grounding for normative truths, because normative discourse enjoys a certain autonomy in relation to naturalistic forms of discourse; see Scanlon (2014), Dworkin (2011), and, for a general version of the strategy, Price (1988). See, too, Gert (2012). This position is bound to be attractive for anyone who despairs of finding a naturalistic account of ethics, but it is hardly going to appeal otherwise. Rejecting such despair, I do not consider it here.

(3.) The focus has usually been on the expressivist reading of judgments of desirability, but a number of authors have also sought to develop a similar reading of responsibility ascriptions. See, for example, Hart (1948–49).

(4.) The intuitive notion of grounding presupposed in the text is meant to fit with the more technical notion employed in recent literature. See Rosen (2010) and Fine (2012).

(5.) If ordinary usage leaves open the possibility of interpreting the concept in one way or another, as it often will, then it is necessary to decide between the candidates on the basis of which property fits best in our theory of the world; for

naturalists, a consideration in favor of one candidate rather than another will be that the grounding conditions involve only naturalistic properties.

(6.) For this account of reduction see Jackson (1998); see also Jackson and Pettit (1990) and Chalmers and Jackson (2001). For a general account of the methodology, with a comparison to the genealogical method adopted here, see Pettit (2019a).

(7.) For the materials needed in a full explication of this functionalist sort of analysis, see Stalnaker (1984) and Pettit (1998)

(8.) Many varieties of a doctrine nowadays known as constitutivism are special forms of reductivism in our sense, although not necessarily naturalistic reductivism. For an overview see Smith (2017).

(9.) In more traditional functionalist terminology, the candidate property, P, that is selected from among multiple candidates, may be treated as the best deserver as distinct from a possible deserver of the name or term at issue: say ‘C’. At least that will be so, if selecting that candidate is identified with deciding on the best account of the property, P (Rosen 2010). The exercise of analysis in such a case narrows the field of potential accounts, ensuring that whatever account is adopted, it will identify a property intuitively ascribable by ‘C’; it will not change the subject. See Pettit (2019a).

(10.) For an example of varying a genealogy in order to make better sense, allegedly, of a given concept, consider the variation to Hart’s genealogy of law proposed in Pettit (2015c, 105–106). This would introduce a development under which it becomes a matter of common awareness that anyone can presume to speak for the group, without fear of contradiction, when arguing, in defense of the law, that it articulates accepted standards as to how things are done in the society. The variation is designed to make sense of why law is taken, allegedly, to be a system that enjoys widespread endorsement.

(11.) This advantage is related to an attractive feature of reconstructive analysis that is not of much relevance in the current context. The approach does not presuppose the availability of a language in which to analyze the target concept. Thus, it is possible to employ reconstructive analysis in making sense of how we use the semantically most basic terms in our language—bedrock concepts, as David Chalmers (2011) calls them—in a rule-following way. See Pettit (2002, Pt 1).

(12.) This is not to suggest that reductive analysis is impossible. For an analysis that fits well with the genealogy developed here, see Jackson and Pettit (1995) and Jackson (1998).

(13.) This charge is laid against a range of studies in De Scioli and Kurzban (2009, 2013). For an exception, see Kitcher (2011, 412), who describes ethics as "an evolving practice, founded on limited altruistic dispositions that were effectively expanded by activities of rule giving and governance."

(14.) For an overview of some currently influential approaches in broadly this category, see Tomasello (2016, 137–143). He divides them into approaches that primarily concentrate on the role of sympathy and reciprocity, on the independent aspects of our psychology that are activated in moral concerns, and on the role of cultural selection.

(15.) Rousseau (1997, 132) seems to have thought in this way about his project in the *Second Discourse*: "The Inquiries that may be pursued regarding this Subject ought not be taken for historical truths, but only for hypothetical and conditional reasonings; better suited to elucidate the Nature of things than to show their genuine origin." I am grateful to Alison McQueen for drawing my attention to this. For the various strands in Rousseau's genealogy, see Neuhouser (2015).

(16.) The genealogy pursued here, as mentioned in the text, aims at explaining ethics, not explaining it away; in that respect it contrasts with the debunking genealogy of ethics offered in the nineteenth century by Nietzsche (1997); see Williams (2002) and Prescott-Couch (2014). Insofar as it is counterfactual rather than historical, it fits with some recent uses of the term *genealogy*, but not with all. See Williams (2000); Craig (2007); Skinner (2009).

(17.) The current text takes the notion of reliance that it employs as fairly intuitive. For a useful analysis, see Alonso 2014.

(18.) On the notion of joint intention and action see (Bratman 2014). For other approaches to the analysis of this notion, any one of which would work for our purposes here, see Tuomela (2007); Searle (2010); Gilbert (2015).

(19.) For a theory of reasoning of broadly the sort that is sketched here, see Broome (2013) and Pettit (1993, 2007b); and for a summary of the main points of commonality, see Pettit (2016a).

(20.) For a naturalistic account of rule-following, see Pettit (1993, 2002, 2007a). With the rules envisaged, there is a further question as to whether at base they are broadly representational rules, as I assume, or inferential rules, as Robert Brandom (1994) and Ralph Wedgwood (2007) argue. The argument developed here is independent of the answer given to this question.

(21.) Suppose that from the earliest stages our protagonists in Erewhon would have been in a position to spell out a notion of epistemological desirability. That concession would not be destructive of the approach taken, since there is no

clear path from epistemological to moral desirability. This remark is made in the spirit of a challenge. It invites those who disagree, and in particular those who reject the sort of genealogy developed here, to tell a purely epistemological story about how we might have become an ethical species.

(22.) If the two narratives identified the same set of properties as the referents of prescriptive terms, of course, that would increase the plausibility of those candidates; it would suggest that they were doubly salient both for the protagonists in the narrative and for us.

(23.) In the later exchange with Michael Tomasello, I consider and reject an alternative, presumably more parsimonious story, in which language is not necessary for morality.

(24.) But, to anticipate chapter 5, even the members of Erewhon might valorize a corresponding desideratum—that of people being helpful and cooperative with one another—when thinking, as they are bound in part to do, from a common point of view.

(25.) Consistently with the line taken in this study, it is possible for two societies to differ in the desiderata that get identified as robustly attractive enough to sustain avowed desires and intentions and, to anticipate, as candidates for making various scenarios desirable. But, assuming they regard one another as conversable, to anticipate an idea introduced later, the members of the different societies may work at achieving mutual understanding; may be able to see why they are moved on each side by the locally favored desiderata; and may come to see those differences, like the differences that may exist within their own ranks, as consistent with realism (Wiggins 1987). For a contrary point of view, see Harman (2000).

(26.) Beginning from a situation of inequality, Nietzsche (1997) explains the emergence of an egalitarian sort of morality, whereby the weak hold back the strong, but the explanation is debunking in the sense that it makes ethics look unappealing rather than appealing.

(27.) Assuming that it does not depend essentially on such suppression does not mean denying that suppression occurs: domination will certainly materialize there (Pettit 2014b), no doubt also hermeneutic injustice (Fricker 2007).

(28.) The set of concepts employed in an elite code that presupposed the suppression of others would be unlikely to count as expressively equivalent to our ethical concepts; and the associated, debunking story would be unlikely to count as a story about the emergence of ethics, as we conceive of ethics.

(29.) For a seminal work on explaining the emergence of cooperation, which had an influence across a range of disciplines, see Axelrod (1984). And for work

pursued in a similar vein, see Pettit (1986; Pettit and Sugden (1989); Pettit (1990); Brennan and Pettit (2004).

(30.) A well-known approach that approximates this sort of narrative, although it also has elements of an emergence narrative, is provided by David Gauthier (1986).