CHAPTER 2

Normative Guidance

§10. The Limits of Altruism

Imagine a population of organisms with altruistic dispositions. For each of these organisms, there is a variety of contexts and a range of other members of the population such that the psychological states of the focal organism—specifically the desires and the emotions—will adjust to reflect that organism's perceptions of the wants, needs, and feelings of the others. These dispositions enable the organisms to function as a population, to live in the same place at the same time and to encounter one another daily without too high an incidence of social friction and violence. But the dispositions are limited: cooperators are sometimes exploited, returns are uneven, and, when there is an opportunity for large selfish benefits, even long-standing allies are sometimes left in the lurch. Defections threaten to tear the social fabric, and, in their wake, much signaling is required; our organisms engage in prolonged bouts of mutual grooming and other forms of physical reassurance.

I shall call these organisms "hominids," although it would be equally apt to dub them "chimpanzees." The limitations of their psychological altruism cause the tensions of their social lives and prevent them from gathering in much larger groups and participating in more complex cooperative projects. A look at their evolved descendants some quarter of a million generations later discloses that the limits have been transcended. Ten thousand years before the present, those descendants have formed settlements that sometimes contain a far larger population; they have learned to interact peacefully with many conspecifics whom they do not encounter on a daily basis; and they have constructed complex systems of cooperation that involve marked differentiation of roles. How has all this been achieved?

One possibility is that they have acquired some new and stronger mechanism for psychological altruism. Conceiving hominid societies as exactly like those of contemporary chimpanzees (or bonobos) is plainly implausible, for the members of the later hominid societies had diverged from their evolutionary cousins five million (or more) years ago. Perhaps as hominid brain size increased, it was necessary for babies to be born at developmentally earlier stages (so their heads would still pass through the birth canal), with the consequence that they were more dependent for a longer period of time. The resulting selection pressure may have favored enhanced altruistic tendencies in the specific context of providing care for helpless young.[1] Without underestimating the importance of steps like these, it is evident that neither hominids nor contemporary human beings have escaped entirely from the difficulties and tensions, the rivalries and conflicts, of chimpanzee social life. If human societies are less vulnerable to breakdown than those of our primate relatives, it is because other modifications have taken place.

What modifications? The hypothesis to be explored is that other changes that have occurred, including in particular the acquisition of language, have made it possible for human beings to reinforce the original limited altruistic tendencies, so members of human societies no longer falter quite as frequently in their cooperation. Because defection is

[1] Arguments along these lines have been developed by Kristen Hawkes and her colleagues (see, for example, Hawkes, James O'Connell, Nicholas Blurton Jones, Helen Alvarez, and Eric Charnov, "The Grandmother Hypothesis and Human Evolution," in *Evolutionary Anthropology and Human Social Behavior: Twenty Years Later,* ed. L. Cronk, N. Chagnon, and W. Irons [New York: De Gruyter, 1999]) and by Sarah Hrdy, *Mother Nature* (New York: Pantheon, 1999), and *Mothers and Others* (Cambridge, MA: Harvard University Press, 2009).

more often prevented, less time has to be spent in reknitting the social fabric. The cumbersome peacemaking of our original hominids is replaced by a new device, one preempting rupture rather than reacting to it, and in principle capable of operating in a wide variety of contexts.[2] That device is necessary for what we think of as ethical practice. I shall call it a "capacity for normative guidance."

The previous chapter was at pains to defend attributions of psychological altruism and to rebut the skeptical insistence that sees Machiavellian intelligence behind apparently helpful or kindly actions. Its account, however, was entirely consistent with the thesis that the psychological altruism of our hominid ancestors was limited. Recall two of the dimensions of altruism: range and scope. An animal may be disposed to respond altruistically to particular other members of its social group ("close friends") across a relatively broad set of contexts, and to respond to all members of its social group in some contexts (banding together against outsiders, for example), although there are occasions on which it would act selfishly even toward its closest friends and staunchest allies. The limited quality of chimpanzee-hominid altruism, in both range and scope, set the stage for the emergence of normative guidance.

The limits of altruism are most starkly and spectacularly visible when the selfish rewards for deserting erstwhile allies are extremely high—as when a male has the opportunity to achieve dominance in the social group. A study of "chimpanzee politics" in the colony at Arnhem (an environment allowing the animals to retain important features of their life in the wild, but, at the same time, providing opportunities for systematic observation of them) revealed the ways in which three high-status males—Yeroen, Luit, and Nikkie—related to one another and to the high-status females, during times of transitions in power.[3] Each male exhibited social behavior readily interpretable as aimed at retaining dominance, achieving dominance, or, at worst, serving as the principal lieutenant of the dominant male. In the early phases of the struggle, Luit aided the newly

[2] It is thus more than a special-purpose mechanism, like the hypothetical emotional disposition that underlies alloparenting.

[3] Frans de Waal, *Chimpanzee Politics* (Baltimore: Johns Hopkins University Press, 1984).

adult Nikkie in achieving dominance over the females, while Nikkie's diversionary tactics enabled Luit to dethrone the previously dominant Yeroen. Once he had attained alpha rank, Luit's policy changed. He consolidated his position by siding with the females and with Yeroen against Nikkie. The abruptness and decisiveness of the switch can easily inspire the conclusion that chimpanzee politics is thoroughly Machiavellian. Apparently, friendships among chimpanzees are situation linked.[4] The subsequent twists and turns of the story seem to underscore that judgment. Yeroen deserted Luit to form a coalition with Nikkie, so that Nikkie eventually became dominant with Yeroen as his lieutenant. After a subsequent period of tension between the two allies, Luit reemerged at the top of the hierarchy, apparently in a weak coalition with Nikkie. The uneasy situation was ended by a night fight, in which Luit was fatally injured by the other two.[5]

To say there are no stable friendships within chimpanzee communities is too strong, for some alliances endure for years, even for virtually the entire lifetimes of the animals—as §9 insisted, the friends of one's vulnerable youth are often one's lifelong companions.[6] Moreover, before the political instabilities in the Arnhem colony, Yeroen and Luit had been longtime allies. The fascinating (but sad) story of the months of conflict reveals—as do similar examples, less fully documented, among wild chimpanzees—how the presence of a clear opportunity for self-advancement can expose the limits of altruistic dispositions. Observers

[4] De Waal offered a sober judgment of the relationships he observed: "Coalitions based on personal affinities should be relatively stable; mutual trust and sympathy do not appear or disappear overnight. . . . If friendship is so flexible that it can be adapted to a situation at will, a better name for it would be opportunism" (*Chimpanzee Politics*, 128). Readers of de Waal's subsequent books (*Peacemaking Among Primates* [Cambridge, MA: Harvard University Press, 1989]; *Good Natured* [Cambridge, MA: Harvard University Press, 1995]; *Primates and Philosophers* [Princeton, NJ: Princeton University Press, 2006]) may be surprised by his early emphasis on hard calculation—for the later work is softer in tone and more inclined to highlight the "good-natured" aspects of primate behavior. The account I offer in the text supplies a perspective from which all of his evaluations can be endorsed.

[5] De Waal, *Peacemaking Among Primates*, chap. 2. De Waal makes the important point that Luit's desire to remain with his social group was so strong that it was difficult to remove him, even after he had been severely wounded.

[6] See Jane Goodall, *The Chimpanzees of Gombe* (Cambridge, MA: Harvard University Press, 1986), chap. 8.

have seen enough varied contexts in which two animals respond to each other to assign them to each other's range of altruism—until the animals encounter a new type of context, in which an altruistic response would require the forgoing of huge potential gains. The selfish action in that context is a sign not that everything in the past has been opportunism, but just that the altruistic disposition is incompletely pervasive. Even for animals who are central to the *range* of the altruist's altruism, there are circumstances outside the *scope* of that altruism.

The conception of psychological altruism offered in §§3–5 reveals what is occurring. Chimpanzees (and our hominid ancestors) have regular psychological propensities for making an altruistic response to another member of their group, with the intensity dependent on salient features of the circumstances. Even though an animal frequently displays a tendency to accommodate the wishes and needs of a particular band member—a "friend"—there are environments in which the intensity of the altruistic response drops to zero. In those environments, altruism suddenly vanishes. Friendship is "situation linked" because there is no fixed value of the intensity of the altruistic response depending solely on the strength of the relationship.[7] Even in the most committed mutually altruistic relationships, circumstances offering one party the chance of a huge advantage diminish the intensity of the response. When the stakes are high enough, it disappears entirely.

The struggle for dominance presents in high relief contours visible in more mundane settings. Every day in chimpanzee troops, members who are not one another's principal allies act in blithe indifference to their fellows' obvious plans. Attempts to obtain a valued object are blocked or thwarted, requests to share food are turned down, appeals for aid in conflict are ignored. The animals involved are not entirely indifferent to one another, for they would band and bond together in the face of an externally presented threat. Rather, the scope of their mutual altruism is

[7] The approach adopted here has some kinship with Walter Mischel's emphasis on the failure of cross-situational consistency in people who have stable personality profiles. See W. Mischel and Y. Shoda, "A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure," *Psychological Review* 102 (1995): 246–68. I am grateful to George Mandler for the suggestion that I explore Mischel's work.

very limited; only under the most threatening situations is it exercised. For the rest, they operate on the basis of tolerance of one another's presence, though, when one's indifferent course collides too strongly with the plans of the other, conflict may erupt.[8]

The limitations of psychological altruism thus show up both in the breakdown of close ties under special conditions (the Yeroen-Luit-Nikkie saga) and in the everyday frictions of animals whose altruism toward one another is limited in scope. The bounds of altruism are revealed in a third way. Even when an altruistic response is made, and when it directs a helpful action toward another animal, there are sometimes signs of psychological division. Conflict *within* is occasionally visible. Chimpanzees are openly torn between selfish and altruistic courses of action, making it apt to attribute to them *two* desires, both expressed in facets of their behavior. An animal hesitates. Holding a branch rich in leaves, he is poised to strip them off and eat, and, simultaneously, the set of the body acknowledges the presence of an ally; eventually, the arm is extended, thrusting a small bunch of leaves toward the friend, while the rigidity of the gesture and the averted face show the presence of a contrary desire. The configuration of limbs and muscles is genuinely a mixture. The tension of the moment is apparent.[9]

Animals can have stable dispositions to respond with quite different intensities of altruism to environmental cues that can simultaneously be present. Some features of recurring situations trigger an altruistic response at a particular intensity; in response to different features the animal is disposed to react with a different intensity of altruism, or perhaps to react with zero intensity. No conflict appears until the animal encounters circumstances with both sets of features: the begging gesture elicits the disposition to share; the lushness of the leaves excites the tendency to consume. The conflict may be resolved through an action expressing only one of the desires, or there can be a compromise, a minimal sharing, or the muscular tension expressing a psychological struggle.

[8] In terms of the discussion of §9, the extremely limited altruism profiles displayed in such cases express membership in different—and often competing—subcoalitions.

[9] Because it is so banal, this phenomenon is rarely described in studies of chimpanzees. Even a few hours of observation will provide instances.

Human behavior reveals similar phenomena. People trying to lose weight are tempted by the aromas from the kitchen. They describe themselves both as wanting and as not wanting the food, and the incompatible wishes are expressed in the active salivary glands and the hasty retreat. Although there is a philosophical temptation to tidy up such cases, to discover a single preference capturing what the person "really" wants, there are, as with the chimpanzees, examples challenging the idea of a single consistent disposition. People who struggle to master a new language or to set themselves a regular regime of exercise can, with equal justice, sometimes be seen as either weak in resolve or healthily unwilling to drive themselves. To find a "real self" free from conflict, we should have to decide which of the candidates is Jekyll and which is Hyde.[10]

If the altruistic dispositions of chimpanzees (and hominids) were limited in the three ways I have described (through breakdown of the most intense responses in extraordinary situations, through the everyday frictions of more casual friends, and through internal conflicts), their social lives would be very difficult. They are (and were). Peace and mutual tolerance are typically hard-won. Precisely because of this, observations of chimpanzee societies disclose periods of intense social interaction, lengthy bouts of grooming undertaken to reassure friends who have been disappointed by recent behavior. At times of great tension within a group, chimpanzees can spend up to six hours a day huddled together, vastly longer than any hygienic purpose demands. Even when daily life is relatively smooth, the minor difficulties and irritations stemming from the incompleteness of altruism, specifically the indifference to one another of animals who belong to different subcoalitions, require an expenditure of time and effort in mutual reassurance. Psychologically altruistic dispositions make it possible for these animals to live together, but the limitations of those dispositions subject their social lives to strain. Day after day, the social fabric is torn and has to be mended by hours of peacemaking.

[10] I draw the examples considered here, as well as the helpful Jekyll-Hyde metaphor, from Thomas Schelling's valuable discussion in *Choice and Consequence* (Cambridge, MA: Harvard University Press, 1984), particularly chap. 3, "The Intimate Contest for Self-Command."

Once, that was the predicament of our ancestors, too.[11] They overcame it through acquiring a mechanism for the reinforcement and reshaping of altruistic dispositions, and for the resolution of conflict. The evolution of that mechanism, the capacity for normative guidance, was an important step in the transition from hominids to human beings.

§11. Following Orders

An ability to apprehend and obey commands changed the preferences and intentions of some ancestral hominids, leading them to act in greater harmony with their fellows and thus creating a more smoothly cooperative society.[12] A capacity for following orders can be expressed in all sorts of actions, many of which have nothing directly to do with making up for the limitations of altruism. Self-command, a familiar human capacity, can address the kinds of problems just discussed.

Those problems, *altruism failures,* are constituted by occasions on which an animal *A,* belonging to the same social group as an animal *B* toward whom *A* is in other contexts inclined to make an altruistic response, fails to respond altruistically to *B,* either forming no altruistic preference at all or acting on the basis of a selfish desire that overrides whatever altruistic wishes are present. The simplest—and original—form of normative guidance consists in an ability to transform a situation that would otherwise have been an altruism failure, by means of a commitment to following a rule: you obey the command to give weight to the wishes of the other. *A* and *B* belong to the same social group, and, for a range of contexts *R, A* forms preferences meeting the conditions on psychological altruism (the conditions of §3). Under circumstances *C,* however, *A* does not respond altruistically to *B* but retains the desire present in *C**,* the solitary counterpart of *C* (or, for examples of internal conflict, it is this selfish desire that leads *A* to action). Under normative guidance,

[11] If our hominid ancestors lived in societies more akin to those of contemporary bonobos, then their situation would have been less tense than under a chimpanzee form of sociality. The differences, however, are matters of degree, not of kind.

[12] Eventually it also modified our ancestors' emotional lives.

A obeys a command that enjoins behavioral altruism: *A* is to act in the way a psychological altruist would; that is, the desire expressed in the action is more closely aligned with *B*'s wishes than the selfish desire would have been.

Just as the discussion of psychological altruism began from a special example (the sharing of food), so here too a particular case is helpful; complications come later. Imagine two members of the same social group, *A* and *B*. They share with each other across a wide variety of circumstances. Faced with an extremely rich and attractive food item, however, *A* is not disposed to form the altruistic preference generated in other sharing situations; the intensity of *A*'s altruistic response vanishes entirely. (In terms of the averaging model, although *A* sometimes sets the value of w_{Alt} at a value greater than 0, under this particular circumstance, *C*, the value of w_{Alt} is 0.) If *A* is now capable of normative guidance, and if the normative guidance takes the very special form of *A*'s commitment to a command that orders food sharing in *C* (perhaps it is the command: "Always share equally with *B*!"), then the preference *A* forms in *C* will take *B*'s wishes into account, by setting $w_{Alt} > 0$ (if the command enjoins equal sharing, $w_{Alt} = 1/2$).[13] If the preference formed leads to action, *A* no longer commits an altruism failure but is behaviorally altruistic. The newly formed desire satisfies conditions 1 and 2 of the account of psychological altruism (§3), but not necessarily conditions 3 and 4. *A,* following orders, need not be responding to any perception of *B*'s wants, nor need *A* be free of the taint of Machiavellianism. Normative guidance transforms the animal's psychological life so that *something that looks, from the outside, like an altruistic preference* is formed (or is operative) across a broader range of contexts.

Psychological altruism was characterized in terms of the difference made to one's own wishes by the perceived presence (and needs) of others; now normative guidance is conceived in terms of the difference made to one's action-guiding preferences by the recognition of

[13] In the discussion of psychological altruism, where *A*'s own perspective is crucial to the formation of the altruistic preference, I saw that preference as incorporating *A*'s perception of *B*'s wants. Here I imagine the command as requiring alignment with *B*'s actual wants. There will be no discrepancy, when *A* has an accurate perception, and, for the time being, I shall assume that mistakes are not made.

commands.[14] The modified preferences, however, need not be fully psychologically altruistic—they just are different from the blatantly selfish wishes that would have prevailed in their absence. The critical idea is the replacement of a desire that fails to incorporate the perceived wants of another individual with an action-guiding desire that gives the other's preferences some weight. That can be achieved even though the desire is not generated by the perception of the wishes of another, and even though it violates the anti-Machiavelli condition. Behavioral altruism (directed by preferences modified so they are closer to the wants of the beneficiary) will sometimes do.

Normative guidance produces surrogates for psychological altruism in animals who can follow orders. The *products* of normative guidance (in its simplest and original form) are desires that issue in behavioral altruism. To understand the *process* of normative guidance, the following of orders that replaces altruism failure by behavioral altruism, it is necessary to probe psychological causes more thoroughly than has yet been done, both with respect to the lives of normatively guided individuals and with respect to psychological altruists. For it is tempting to adopt an oversimplified (and overly neat) picture of the distinction between normative guidance and the mechanisms behind full psychological altruism.

On this oversimple view, psychological altruism is generated by an *emotional* response to the beneficiary, whereas normative guidance involves the operation of a *cognitive* faculty ("reason," perhaps). Psychological altruism is "hot," normative guidance "cold." Both subtheses should be rejected. Start with the varieties of psychological altruism.

Different kinds of psychologically altruistic individuals are possible. Imagine an altruist who reacts in context *C* by modifying his or her wishes from those occurring in the solitary counterpart *C*,* because of his or her perception of the wishes of *B;* the new desire may be accompanied by the presence of an emotion toward *B,* and, if present, the emotion may or may not cause the new preference. Even if we use a crude and unana-

[14] Plainly, one can recognize commands and act in response to them in ways that have nothing to do with psychological altruism. That will concern us later. For the time being, normative guidance is tied directly to the reshaping of altruism.

lyzed concept of emotion to consider the situation, we can distinguish four cases:

a. *A*'s new desires are caused by an emotional response to *B*.
b. *A's* new desires are not caused by, or accompanied by, any emotional response to *B*.
c. *A's* new desires are not caused by any emotional response to *B,* but the factors that generate the new desires also produce in *A* an emotional response toward *B*.
d. *A*'s new desires are not caused by any emotional response to *B;* an emotional response to *B* accompanies those desires, but it is independent of the causal process that generates the new desires.

The oversimple view supposes that cases of type a represent the most fundamental (primitive) form of psychological altruism; cases b–d display responses that could emerge only from normative guidance.

Why should one think this? Underlying the view is an apparently plausible line of argument: the adjustment of desire could result only from the operation of an emotion or the outcome of a process of reasoning; prior to the articulation of ethical practice, the only forms of reasoning available to an agent (human or nonhuman) would have to be calculations of selfish advantage; hence, preethical adjustments of desire based on reasoning would fail the anti-Machiavelli condition; by the same token, the only ways in which obeying commands could produce altruism involve the recognition of reasons for modifying desire.

On the account of §3, all four types count as instances of psychological altruism. The argument just outlined denies that the modification of desire constitutive of *psychological* altruism could occur in cases b–d. To assess it, consider the examples that occupied us in the last chapter. Some of them fit easily into the simple view. Prominent instances of psychological altruism among primates express an emotional reaction to the plight of another animal: mothers' immediate responses to the discomfort of the young, or Little Bee's patience with her mother. It is far from evident, however, that the example of Jakie and Krom can be so easily assimilated. Further, as §6 argued, maternal concern is not always a matter of being prompted by emotion. The primate mother who stumbles across a carcass and views it as an occasion for seeking out her young appears to be undergoing more complicated psychological processes, which are not easily captured in a venerable—but crude—philosophical practice of opposing reason to the passions.

On the ecumenical view adopted in §4, emotions are complex processes typically involving both cognitive and affective states. The causal relations among these states can be quite various, and there is no reason to suppose that the cognition cannot be primary. Perhaps a cognition—recognizing that Krom wants the tire and that she has failed to remove it, seeing that this carcass is food for the young—induces a new affective state. Or perhaps that cognition leaves the prior affective condition of the perceiver unaltered—there is no upsurge of emotion at all, but simply the formation of a new desire on the basis of affective dispositions already present. Animals can have propensities for forming new desires that do not depend on their entering into a new affective state. Consequently, versions of b–d can count as psychological altruism.

Not only can cognition cause affective states, or produce new desires without modifying the affective background, but there can also be intricate chains of causation in which perceptions give rise to new beliefs, the new beliefs generate affective states, these affective states, in turn, lead to altered beliefs, the altered beliefs to novel affective states, all this entangled with the formation of desires: indeed, this may be the stuff of much of our more complex emotional life. The simple vocabulary employed in the examples a–d is inadequate to present clearly all the ways in which psychological altruism can arise (even though we do not yet know just what form a fully satisfactory conceptualization of the emotions would take). Moreover, there is no basis for denying at least some of the complex possibilities to nonhuman animals.

This brief for taking the complexities of our emotional life seriously subverts one-half of the simple view. Troubles also beset the other half, the proposal that normative guidance must be a matter of reasoning. Recent work in neuropsychology suggests that the opposition of "cold" reason to ardent passion is highly problematic and that there is evidence for the role of emotion in what have often been viewed as cool deliberations.[15] Beyond this general point, there are grounds for attributing a major directive role to emotions in some instances of normative guidance.

Consider, first, the way in which the psychology of a normatively guided individual can develop. Initially, a human being, a member of one of those small bands in which our ancestors lived, is disinclined to respond to the predicament of one of his fellows. Capable of normative guidance, he obeys a command to make a behaviorally altruistic response, and his reacting in this way generates in him an emotional response to the beneficiary, a primitive feeling of sympathy (as in case c previously). That feeling is reinforced by the beneficiary's reaction to his behavior, and, perhaps after a few further interactions, this person is able to engage in the behaviorally altruistic conduct *either* on the basis of the original process *or* through a full—psychologically altruistic—identification with the other. An emotional change may thus be the direct product of the commitment to following an imperative: as you come to endorse the command to treat your brother in a particular way, your emotions toward the brother are modified, and the new fraternal feeling gives rise to the desire to treat him in ways you would previously have avoided (or resisted). Initially, normative guidance operates to produce behavioral altruism, but it eventually issues in full psychological altruism.

How is that first step taken? Must it be on the basis of reasoning—perhaps through a Machiavellian recognition of the benefits of complying? Not necessarily. Endorsing the command can embody emotions, sometimes emotions directed toward the commander: you may accept it because you are afraid.

The point may provoke an obvious reaction. If the notion of normative guidance is liberal enough to allow for conformity grounded in fear, acquisition of the capacity for normative guidance cannot be the decisive transition to ethical practice. A dilemma seems to loom. If the ability to follow commands, to obey rules and precepts, is the decisive step in acquiring a *genuinely ethical* practice, then this special sort of ability

[15] See Antonio Damasio, *Descartes' Error* (New York: Putnam, 1994), and Marc Hauser, *Moral Minds* (New York: Ecco, 2006).

requires an explanation—for it cannot be rooted in emotions of fear or prudential calculation. On the other hand, if processes in which people comply because they fear the consequences of disobedience were available to our human ancestors and initiated the practice of normative guidance, then only a *simulacrum* of ethical practice has been connected with the prior preethical state; the people in question have not yet made the transition to the *real thing.* These individuals, allegedly "subject to normative guidance," have not yet achieved the distinctively "ethical point of view." The broad conception of normative guidance allows for an evolutionary transition from hominids lacking the capacity to humans who enjoy it, but this continuity is purchased at the cost of losing contact with the proposed goal, to wit, the emergence of ethics. To make normative guidance relevant to ethics, one needs a propensity to act in accordance with commands grounded in a different (and purer) form of psychological causation.

There is no such purer form to be had. At least since the eighteenth century, philosophers who have disputed the character of ethical agents have envisaged an "ethical point of view" in which people give themselves commands—commands that are not external but somehow their own, the "moral law within"—and have regarded this point of view as requiring the subordination, if not the elimination, of emotion.[16] Others have regarded the operation of emotion as central to ethical agency. It is often assumed that the major challenge for a naturalistic approach to ethics consists in showing how the achievement of the "ethical point of view" might have evolved from more primitive capacities; inspired by this thought, naturalistically inclined thinkers frequently address the challenge by attempting a reduction of that "point of view" to the feeling of special types of emotions. Their disputes with their opponents rest on a shared mistake.

The acquisition of a capacity for normative guidance—understood, as above, as an ability to follow orders that issues in surrogates for altruism—

[16] A prime source of this view is, of course, Kant, and the most sophisticated elaborations of it are offered in the Kantian tradition of ethical theory. Yet Kant's opponents, who often protest the denigration of the emotions, share the emphasis on a distinctively ethical point of view. I am proposing that we reject a precondition of their debates.

does not mark the transition to the "ethical point of view." That is not because there is some further move that does the trick awaited by the critics, one that shows how a very special kind of normative guidance (a special way of internalizing the orders, say) constitutes the "ethical point of view," *but because the entire conception of the "ethical point of view" is a psychological myth devised by philosophers.* There are plenty of ways in which human beings can be led to recognize and to conform to commands. While it is undeniable that some kinds of causal processes make ethical *progress* over others (in ways Chapter 6 explores), we should not infer a binary distinction between those processes that constitute genuinely ethical motivations and those that do not.

Most of the people who have ever lived have embedded their ethical practices in a body of religious doctrine, viewing the precepts to be followed as expressions of the will of gods, spirits, or ancestors (or occasionally as capturing the tendencies of impersonal forces). Fear, awe, and reverence have been parts of the emotional backdrop to most of the important decisions and deliberations these people have made, and virtually all those decisions have been subject to felt concerns about the attitudes of transcendent beings. The fact that these people have presupposed massively false beliefs about the universe does not undermine their status as ethical agents. Neither should the fact that what they want, intend, and do are partially caused by emotions of fear and awe. To insist on an "ethical point of view" liberated from such emotions is to reserve that point of view for a very small number of cool secularists. Moreover, it is reasonable to worry that the alleged ethical point of view is itself only available because of the perspectives previously adopted by those no longer counted as full ethical agents. The ability to "revere the moral law" probably depends, in the evolution of culture and in the development of individuals, on prior emotions, simpler feelings of reverence now written off as ethically primitive.

There are many different ways in which people can be led to behavioral altruism through their commitment to obeying a command. They may explicitly represent to themselves the consequences of disobeying, and find those consequences unpleasant or frightening because of future interference with their bodies, behavior, or projects. They may make no such explicit representation, but be moved by fear, or respect for the commander. They may regard the source of the command as a being greater than themselves, one whom it is important to obey. They may actively want to be in harmony with the wishes of some such being. They may regard the source of the command as part of themselves, and fear the psychic disharmony caused by disobeying it. They may have a general idea of the worth of the situations brought about by commands of a general type to which this particular imperative belongs. They may want to be the sort of person who lives in accordance with a general class of commands they have previously endorsed. They may want to live in harmony with others who expect that commands of this sort will be obeyed. They may have a general ideal for themselves that involves obeying commands current in their social group. Or they may conceive of themselves as members of a joint project, in which commands are issued and obeyed. These surely do not exhaust the possibilities, and some of the considerations can be present together, with different degrees of force.

The merits of a liberal articulation of the concept of normative guidance should now be apparent. Our decisions involve a hodgepodge of considerations and feelings, and it is foolish and unnecessary to limit the full range of psychological possibilities, taking some to be importantly free of emotion and others not, some to be constitutive of "the ethical point of view" and others not, some to accord with the anti-Machiavelli condition and others not. Emotions are complex processes typically involving both cognitive and affective states (§4), causation can run from affect to cognition or in the opposite direction, and our actions sometimes result from intricate cycles involving different types of states. The simple view, against which I have been campaigning, formulates the possibilities using language we know to be inadequate (even though we surely still lack a clear and precise vocabulary for categorizing the relevant states and processes).

Psychological altruism occurs when perception of the wishes of another modifies desires to align them more closely with the perceived wishes. Normative guidance comes about when the recognition of a command leads someone to act in accordance with it and (in the conditions studied so far, the context of the beginnings of the ethical project) to replace altruism failure with behavioral altruism. Emotions, desires, and cognitive states can be entangled in *both* cases. The causes of psychological altruism and of normative guidance are probably highly heterogeneous. There are many ways to be a psychological altruist and, equally, many ways to undergo normative guidance. None of these latter modes is especially privileged as definitive of an "ethical point of view."

No doubt there are extreme cases. Someone who forms the wish to help another, simply because he is commanded to do so and because he recognizes that disobedience will bring painful punishment on himself, is no psychological altruist and (at best) at a rudimentary stage of ethical practice. At the other extreme, a person who has a general conception of the wishes of others, who follows a rule because it is taken to promote the desires of someone else, may be viewed as at least an approximation to psychological altruism and as participating in a more advanced form of ethical practice, despite the fact that the wishes, and even the situation, of some of those she aids are unknown to her, and even though she has a standing desire to be the sort of person who contributes to the satisfaction of others' desires. Normative guidance, as explicated here, applies to individuals of both types, generating behavioral altruism in the one instance and something akin to full psychological altruism in the other.

Given the diversity of causal possibilities, why would one want to take a stand on which of them has to be realized in a genuinely ethical agent? The "ethical point of view" emerges as a challenge for naturalism because it opposes the idea of ethical agents as those sympathetic individuals who respond to the needs of others. While superficially attractive, these people suffer a defect that makes them less than fully worthy.[17] Their kindly emotions are unreliable: it is reasonable to fear that the mind of "the lover of humanity" will sometimes be "clouded," and that, under

[17] The classic source for the reaction is Kant, *Groundwork of the Metaphysics of Morals*, Mary Gregor, trans., (Cambridge, UK: Cambridge University Press, 1998, Akademie pagination 398). This passage is often viewed as expressing an opposition to Hume, but I suspect that Kant actually had Adam Smith in mind. Not only does Smith develop the notion of sympathy much further than Hume did, but his *Theory of Moral Sentiments* (Knud Haakonssen, ed. Cambridge, UK: Cambridge University Press, 2002) [unlike Hume's *Treatise* (Oxford, UK: Oxford university Press, 1978)] is a work Kant is known to have read.

such conditions, fellow feeling will no longer operate and the person will act selfishly. Yet if we take the concern about reliability seriously, "proper" motivation appears impossible. What basis is there for supposing that carefully restraining the passions and engaging in abstract moral reasoning (of any of the sorts philosophers have commended) will prove reliable? Can't our faculties of reasoning sometimes be "clouded," too? Abstract reflection and reasoning are hardly *more* reliable than the emotional responses dismissed as capricious. Many of the most horrific deeds of the twentieth century were carried out in the name of abstract principles.

As we shall appreciate later, reliability is the issue (§21)—the worry about the "clouding of the emotions" expressed an important point. Yet the search for a single type of psychological causation, invariably reliable or at least always more reliable than its rivals, is foolishly utopian. Different ways of inducing people to modify their preferences and actions through obeying orders have different merits and deficiencies. Normative guidance would work better by taking advantage of the ways in which different psychological processes are suited to different situations. Perhaps normative guidance evolved in parallel fashion to familiar types of organic change, where initially crude systems for producing some important outcome are supplemented with further devices: the organism has a variety of ways of generating what is required and is thus buffered against catastrophe.

Normative guidance almost certainly began with crude external orders, followed out of fear; much normative guidance may have been mediated by respect for the supposed commands of transcendent beings, respect tinged with hopes and fears (§17). Out of those hopes and fears have come quite other emotional resources for motivating obedience, feelings of awe and respect, of social solidarity and of contentment in acting jointly with others, of pride in one's conduct and of responsibility to one's fellows. The history of modes of normative guidance embodies certain kinds of progress, and attempts to act through following dictates the agent sets for himself, considers, and endorses have often been progressive with respect to earlier and cruder forms of psychological causation. These differences, however, are matters of kind rather than of degree. Some processes (perhaps processes involving an especially pure form of emotion, perhaps processes that rein in emotion entirely) are valuable additions to our repertoire, but they have no special standing setting them apart from the modes of normative guidance preceding them. Their merits can be recognized without supposing them to constitute an "ethical point of view," which counts as the last word.[18]

The approach defended here allows a more systematic treatment of the behavior of subjects in economic experiments (§7). These people are recruited by researchers, know little, or even nothing, of one another's wants or needs, and are placed in situations in which they can decide what fraction of a monetary reward to share with fellow participants or how much they will give to punish those who do not act cooperatively. One thing is clear. The participants' preferences cannot be adequately represented by supposing them to be concerned with money and nothing but money: they do not belong to the fictitious species *Homo economicus*.[19] So *why* do they share, or give money to punish? Not because they are moved by the plight of people who would otherwise leave empty-handed, for they lack any basis to make judgments about the impact on these strangers. One explanation, consistent with the evidence, is that some form of normative guidance is playing a role. The participants do what they do, sharing with others, because they follow an order, one they have accepted and endorsed or one they view as current in their society.

If they were genuinely moved by a dedication to fairness, a clear-eyed vision of the value of equality in dividing goods, if this and this alone moved them to want to share (or to punish noncooperators), we might

[18] See Thomas Nagel, *The Last Word* (New York: Oxford University Press, 1997), chap. 6; also *The View from Nowhere* (New York: Oxford University Press, 1986), chap. 9.

[19] This is already to demonstrate something that is very important for economic research, for it entails that models imputing utility functions that are increasing functions of amounts of money, and of this alone, are unlikely to accord with the behavior of actual agents (for whom other things are important). Indeed, for the project of advancing economics, any concerns about the ways in which the subjects come to the wants they express in their actions are entirely irrelevant. What is far less clear is how these ingenious experiments bear on philosophical concerns about altruism and its role in ethical practice. For an illuminating presentation of the experimental work, see Ernst Fehr and Urs Fischbacher, "Human Altruism—Proximate Patterns and Evolutionary Origins," *Analyse & Kritik* 27 (2005): 6–47.

count their preferences as altruistic. Although they know nothing of the needs of those they reward, they have a general view that outcomes in which those people received nothing (or even received less than half) would be, from the perspective of the beneficiaries, unsatisfactory; the sense of fairness endorses the complaint, and so, without any selfish background motive, they want an outcome of equal division. This conjecture might tell the whole truth about some of the experimental subjects, but we are by no means forced to accept it. For the available evidence leaves open alternative modes of normative guidance: perhaps the participants want the "glow" (or to avoid the "pang"); perhaps they want to be the sorts of people who accord with prevalent social norms of sharing; they know their parents, spouses, friends, or children would disapprove of their greedily making off with everything they can; they may want the approval of the experimenter and not want to go down in his or her records (even if only mentally kept) as "one of those stingy people"; without any clear sense of the virtues of equity, they know this is the sort of thing of which people approve, and the sacrifice does not seem too large (they are going to leave the lab with something in their pockets). Elaborated versions of these psychological scenarios raise serious doubts about whether the anti-Machiavelli condition is satisfied. Even more obviously, the modified wants are not responses to *another person;* indeed, in some experiments, the actual beneficiary is invisible; the dialogue is between the agent and the ambient society (perhaps embodied in the experimenter).[20]

Normative guidance can generate full psychological altruism in situations that would otherwise be altruistic failures. Initially, it almost always generates *behavioral* altruism. Human motivation is sufficiently complex that, in many circumstances including those of the economic experiments, we just cannot tell (at least not without a lot of work—and maybe some luck besides) how exactly to classify people who act to benefit others.[21]

[20] Subjects whose primary motivation is to impress (or to avoid disappointing) the experimenter are easily linked to the experimental subjects who were prepared to inflict pain on others.

[21] This conclusion motivates the attitudes of the researchers who carry out these experiments, who suppose the important concept is that of behavioral altruism. My account

§12. Punishment

To treat normative guidance in this way has an obvious presupposition. Behind the disposition to follow orders, whether delivered externally or from internalized commands, must stand practices of punishment. Unless there were sanctions for disobedience, fear could hardly be central to the initial capacity for normative guidance. Conversely, when punishment is present in a group, it can make possible the evolution of elaborate forms of cooperative behavior (and much else besides).[22]

Can this presupposition be defended? The *actual* beginnings of the ethical project have been seen as a transition from a state of limited psychological altruism to one in which commands are followed out of fear. The plausibility of that view would be undermined unless there were an explanation of the *possibility* of punishment.[23]

Begin with chimpanzee societies in which a crude precursor of punishment is already present. Conflicts within these groups are often settled through the interventions of a dominant animal.[24] Here rank or physical strength (or both as concomitants of each other) prevail, and a dispute is settled—not always, of course, through the infliction of pain or discomfort on the animal whose initial defection gave rise to the conflict. Allies who might have intervened to protect some of those who receive the rough discipline of the dominant animal anticipate the costs to themselves and hold back.

of the ethical project also recognizes the important role of dispositions to psychological altruism. Different concepts are needed in different forms of inquiry and there need be no quarrel about which notion of altruism is the "right one."

[22] Here I rely on a brilliant essay by Robert Boyd and Peter Richerson, "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups" (originally published in *Ethology and Sociobiology* 13 [1992]: 171–95; reprinted as Chapter 9 of Boyd and Richerson, *The Origin and Evolution of Cultures* [New York: Oxford University Press, 2005]).

[23] Here, it is important to recall the methodological points of §2. A hypothesis about the actual origins of the ethical project is supported by evidence about the prior hominid state, and recognition of familiar human capacities to address its social difficulties. That hypothesis must be defended by showing that its presuppositions are compatible with the constraints acknowledged by pragmatic naturalism.

[24] Goodall, *Chimpanzees of Gombe*, 321–22; and de Waal, *Peacemaking Among Primates.*

Punishment need not always take so dramatic a form and can be present simply when animals recognize opportunities for cooperation with one another. Once the basic dispositions to altruism toward nonrelatives that underlie chimpanzee-hominid society are present, optional games (§8) are available. There is a pool of potential partners who can be recruited for joint ventures. Because of tendencies to bond with close friends and allies, some kinds of defections in the ventures will be tolerated—animals will not behave with the rigor of discriminating cooperators, refusing invitations to joint activity, when the potential partners are targets of psychological altruism and longtime allies. Nevertheless, as the ties are weaker and the history of interaction more limited, it is to be expected that a strategy like discriminating cooperation will be favored. The altruistic dispositions emerging from the coalition game incline animals to give weight to benefits received by their allies, and thus to increase the value attributed to outcomes in which the ally gains and the focal individual loses; consequently, animals will be less rigorous in dismissing their close friends as potential partners for interaction; as the relationship becomes more distant, however, the deviation from the basic structure of the optional game (for example, optional PD) is much smaller, and the strategy favored will more closely approximate discriminating cooperation, refusing further interaction on the basis of a single defection.

That itself is a form of punishment. To deprive an animal of opportunities for cooperative interaction is to force it sometimes to pursue suboptimal ways of meeting its needs. So long as there are occasions for joint activity with others, allies who remain willing to enter partnerships with the animal in question, the impact need not be severe. If the allies are often unavailable, however, or if the refusal to interact spreads more broadly, life may become quite difficult. Ostracism can be a serious punishment.[25]

The practices just mentioned turn on the responses of individuals toward actions by others, actions they do not like. Those individuals can

[25] Social confinement and exclusion are used as forms of punishment in small human societies. For a vivid depiction of the effects, see Jean Briggs, *Never in Anger* (Cambridge, MA: Harvard University Press, 1970).

effectively cause pain for the perpetrators, either through their strength (or through force that is unchallenged because of considerations of rank) or through refusal to interact (a response even the weak can usually manage). Social participation in these events is minimal: in the one instance, bystanders behave as mere spectators because of the physical power (or the rank) of the punisher; in the other, their attitudes or actions cannot completely undermine the punisher's success—they may continue to cooperate with the animal whom the punisher has blackballed, but they typically cannot compel the punisher to do so.[26] More sophisticated systems of punishment emerge, as animals form social expectations about the circumstances of punishment.

For an action to be a kind, even a crude kind, of punishment, rather than simply another contribution to the melee, it is important that bystanders not be drawn in. Thus, a first step in the direction of punishment requires that other members of the group, even allies of the threatened animal, should not intervene. There is a regularity—friends of the animal(s) targeted in punishment let it proceed. The next stage couples the mere regularity with an expectation, shared across the population, that others will not interfere in such contexts. The expectation suppresses resistance on the part of the target; the animal picked out expects others not to intervene and merely suffers what happens. A further refinement would be the existence of a regularity concerning the animals who carry out the aggression: perhaps they are animals who bear a particular relation to the context; perhaps they play a particular social role. Finally, there arises an expectation about the identities of the animals who initiate aggression. At this last stage, we have reached the systems of punishment found in contemporary human societies (and in societies for which we have historical records).

The actual evolution of punishment may have diverged from the sequence of steps just envisaged; nor is it necessary to specify a point

[26] In principle, just as there could be escalation of violence when some animals physically punish others, so too there could be escalation of noncooperation when a discriminating cooperator crosses another individual off the list of potential partners. In the former case, obvious strength or recognition of rank stops the arms race; in the latter, the refusal of *A* to play optional games with *B* is, I suspect, often not recognized and, when it is, does not inspire *B*'s allies to forgo potentially valuable opportunities for cooperation with *A*.

at which “real” punishment is present; nor has it been explained why any hominid lineage went through these stages. Firm views on the last issue ought to be grounded in precise models of the advantages of moving from one stage to the next, and constructing such models would require far more information than we can probably hope to acquire about the causes of reproductive success in the ancestral environment(s).[27] The challenge is not to understand the *actual* evolution of punishment, but to respond to concern that no such evolution is *possible.* Decomposing punishment into conditions that can be sequentially achieved suffices to demonstrate the possibility of gradual evolution. Crucially, to buttress the account of normative guidance, the emergence of punishment does not require the prior achievement of ethical practice.

The early stages of the envisaged sequence could have originated without language: as noted, chimpanzees sometimes resolve conflict by a crude form of punishment, and the possibility of optional games gives rise to another. By contrast, the later steps would be facilitated by prior acquisition of linguistic skills. The emergence of more sophisticated forms of punishment is probably intertwined with the evolution of language—and both are probably entangled with the acquisition of normative guidance.

Suppose a type of altruism failure, keeping food items for oneself, say, regularly elicits aggressive retaliation from others. Chimpanzees and hominids could recognize the regularity, thus allowing for variants who recognize the potential threats to them if they fail to share, and whose fear generates compliance. With the advent of language, descendants of these variants can formulate the command for themselves and for others. Mothers train their young by commanding them to share, and, because of the command, the young stay out of trouble and avoid risks of injury. The repeated commands leave an echo on later occasions, and the

[27] It is not hard to construct models allowing for the possibility of adaptive advantages in initiating and refining systems of punishment. Those models serve the function of protecting the hypothesis of a gradual evolution of schemes of punishment against the charge that they are idle fantasies, incompatible with Darwinian evolutionary theory. Yet, without far greater knowledge of the ancestral environments, and hence of the values of pertinent parameters, it would be unjustified to propose that any model of this sort picks out the actual course of the evolution of punishment. Modesty is appropriate here.

original disposition to share is reinforced by the memory of maternal instruction.

Through explicit command and fear of punishment, even the primitive punishment of the earliest stages, normative guidance can obtain a purchase. Animals with a capacity for recognizing and following orders have advantages over their fellows who lack that ability.[28] Once the capacity is present, it can operate to yield the socially coordinated behavior required by the more advanced forms of punishment. Animals—now surely human beings—can formulate descriptions of regularities about the consequences of alternative forms of behavior on the part of bystanders. Bystanders who intervene are seen to encounter the same sorts of trouble as the first-order offenders who perpetrate the failures of altruism that invite punishment. Group members formulate, for themselves, their kin, and their friends, orders to stand back and let the discipline proceed. When these rules become prevalent, each can recognize others as complying, yielding a social expectation that bystanders will do no more than watch. Perpetrators, aware of the expectation, see the futility of resistance, commanding for themselves a strategy of docile submission less dangerous than trying to fight back. So normative guidance, once present, can figure in transitions to more refined forms of punishment. As punishment is refined, further regularities become salient, providing scope for additional occasions of normative guidance.

Recognizing the painful consequences of particular—and tempting—courses of action, our ancestors, prompted by fear of the outcomes, ordered themselves (and their offspring) to hold back. The next step will be to consider how the grip of this capacity for self-command and self-control might be intensified.

[28] Once again, whether the capacity will be advantageous turns on the details of the situation. If punishment carries even a small probability of serious damage, and if the order-following variant is just slightly more likely to avoid the altruism failure, then the expected gains in terms of staying intact and healthy can outweigh the loss of food that results from sharing. Once again, we cannot know whether this scenario is plausible; this is a "how possibly" explanation.

§13. Conscience

Two prominent Shakespearean figures present a view of conscience. Richard III offers a conjecture about the origins of internal checks on our conduct:

> Conscience is but a word that cowards use,
> Devised at first to keep the strong in awe.

Hamlet, while using similar words, worries about the effects of conscience on behavior, once the tendency for self-regulation is already present:

> Thus conscience doth make cowards of us all,
> And thus the native hue of resolution
> Is sicklied o'er with the pale cast of thought.

Together, the passages suggest an obvious picture: strong people with self-interested intentions are held in check by an internalized mode of normative guidance that substitutes fear for their "native resolution." That picture has sometimes moved thinkers to lament the crippling effects of internalization.[29] Whether or not they are right, pragmatic naturalism needs an explanation of how internalized commands became *possible.*[30]

The first forms of normative guidance, considered in §11, focused on the capacity to follow explicit orders. Human beings (rather than hominids, since they have acquired language) learned the local rules in childhood and later remembered the commands passed on to them. As they grew in strength, however, the memory of older commands might prove

[29] Nietzsche's complaint is most evident in the first two essays of On *the Genealogy of Morality* (Cambridge, UK: Cambridge University Press, 1994); similar themes are sounded by Freud, in many later works, but especially in *Civilization and Its Discontents* (New York: Norton, 1989), as well as by William James in his writings on the "strenuousness" of the moral life (James "The Moral Philosopher and the Moral Life" in William James *Writings 1878–1899* (New York: Library of America, 595–617).

[30] Once again, the methodological points of §2 are relevant here.

too weak to overlay the "native hue of resolution." They might lapse into the altruism failures from which normative guidance promised liberation.

As more sophisticated systems of punishment are elaborated, however, the ineffectiveness of remembered commands becomes costly both for those who fail to be normatively guided and for other members of their societies. Variant individuals, with a tendency to respond to modes of socialization that reinforced the disposition to self-discipline, would cooperate more thoroughly and encounter less trouble. This extension of normative guidance involves both social innovations and psychological changes in the individuals. On the social side, it requires practices of training the young members of the group so that the prospects of flouting a command become associated with emotions they find unpleasant. On the individual psychological front, it consists in refinements of the emotional lives of these individuals.

The Shakespearean suggestion that fear lies at the root of this process of internalization need not be exclusive: other emotions might be available for recruitment to the cause of normative guidance. Imagine a social group of early humans, able to issue and remember commands, but vulnerable to the flouting of those commands by individuals who think of themselves as strong. An innovation in the training regimes customary among this group, the practice of issuing orders to the young, promotes an enduring fear: perhaps they are lured into violating one of the precepts and then subjected to some extraordinarily harsh and memorable punishment; perhaps this occurs at an especially impressionable age. Thereafter, even as they grow, those trained in this way remain haunted by a sense of dread as they contemplate disobeying certain commands. Conscience does make cowards of them. Yet, similar effects can be achieved in different ways. If the young are induced to identify with some of the orders current in their group, if they see obeying those orders as partly constitutive of belonging to this distinctive social unit, they may feel more complex reactive emotions—pride, perhaps, when they continue to carry out the commands, shame or guilt when they do not. As these reactive feelings attach to outcomes considered in prospect, they may substitute for the raw fear of punishment, promoting the same types of cooperative behavior on a different basis.

We know too little about the intricacies of human emotions to elaborate this scenario in any great detail, but the outline is clear. The simplest modes of internalization trade on the ability of programs of socialization to exploit human fears. More sophisticated methods of training people can foster other emotions, perhaps emotions unavailable in different developmental environments, whose association with potential courses of action reinforces tendencies to behavioral altruism. The result is a society in which cooperation is more broadly achieved and in which costly episodes of punishment are less frequently needed. Further, even at early stages of the ethical project, different groups may have cultivated different emotions, founding their ethical practices in distinctive ways. There may be several ways to build a conscience.

However it is formed, conscience is the internalization of the capacity for following orders. The ably socialized individual does not simply hear the voice of an external commander, or remember the injunctions administered in childhood. The commanding voice seems to come from within, initially and crudely as the expression of fears, later perhaps as the representation of membership in a particular social group. In either mode, it provides a more effective anticipation of the costs of deviating from the approved regularities in conduct than the original tendency to follow and remember external orders. The conscience-ridden human being fits more easily into the social niches, provides less provocation to punishment, and encounters much less trouble.

If, to borrow another phrase from Hamlet, society plays upon the individual as on a pipe, it need not always be the same tune. Successful social inculcation of normative guidance may work through quite different emotional complexes, even though variant group techniques succeed equally in securing cooperative behavior. Although conscience begins in fear, it may later be dominated by shame or guilt, pride or hope, emotions available only in social environments where normative guidance, in some cruder form, has already taken hold.[31]

[31] In accordance with my strategy of *outlining* a scenario, I offer no detailed claims about how any of these emotions is to be understood, or whether, as some anthropologists and philosophers influenced by them have suggested, there are cultures in which the emotion of shame is central and others in which the emotion of guilt is central. As noted in the

Nothing follows about the evaluation of internalized normative guidance. Modes of conscience fueled by fear (or other negative emotions) can surely distort and cripple human psychological lives,[32] but whether self-regulation from internalized fears of authorities must always be so baneful in its effects is by no means clear. The consequences from harmonious interactions with others can outweigh sacrifices in expressing selfish desires—indeed, the social involvement may be viewed as a deeper and more significant articulation of what is properly one's own set of wants and aspirations. Much depends, plainly, on the particular orders that the human with a conscience feels compelled to obey, and whether they interfere with yearnings central to a person's life. There are two dimensions to the internalized forms of normative guidance, one characterized by the emotional basis through which compliance is obtained and one depending on the content of the commands. Repressive forms of conscience can be generated along either dimension, if conscience develops in unhealthy ways. Social inculcation that couples all deliberation to fear, shame, and guilt can warp the socialized individuals; equally, massive prohibitions, however backed by emotional responses, can confine someone completely.[33] On the other hand, a person whose conscience expresses itself in a variety of ways, including sometimes through fear, guilt, and shame, can achieve, and recognize herself as achieving, a richer emotional life through the social exchanges conscientious cooperation promotes.[34]

text, I do not think these exhaust all the possibilities; nor do I think they exclude one another in the ways often suggested.

[32] The point is eloquently expressed by Nietzsche in his critique of the "herd morality" based on *ressentiment*. How to foster forms of conscience that yield the important benefits of internalization without deforming individuals is, of course, a question the ethical project continually has to decide.

[33] One way of reading Freud's *Civilization and Its Discontents* is to view him as claiming that any way of achieving the measure of social cooperation required for civilization will have to involve both prohibitions on a massive scale and pervasive negative emotions. His claims rest on very particular ideas about our fundamental desires and drives.

[34] This is obviously akin to the Hobbesian perspective on the constructive role of fear that permeates *Leviathan*.

§14. Social Embedding

Members of the human groups envisaged (small societies, akin to the hominid bands preceding them) are socially embedded in two important ways. First, as just supposed, the particular way in which normative guidance is internalized depends upon the training regimes present within the group. Second, the content of the orders given depends on discussions among members of the group. The character of the discussions has varied considerably from group to group, time period to time period, with different degrees of involvement according to age, rank, and sex. Originally, however, an agreed-on code, articulated and endorsed after discussions around the campfire,[35] was transmitted to the young through training regimes that had also been socially elaborated and accepted.

Equality, even a commitment to egalitarianism, was important in the earliest phases of the ethical project. In formulating the code, the voices of all adult members of the band needed to be heard: they participated on equal terms. Moreover, no proposal for regulating conduct could be accepted unless all those in the group were satisfied with it.

Although these theses may appear implausibly strong, they rest on three sources of evidence. Anthropological studies of societies whose ways of life are closest to those of our early human ancestors show the types of equality ascribed.[36] Further, if normative guidance is to resolve the social tensions, discussions must end old conflicts, not generate new ones. Lastly, for a small band, one that must work together and unite against external threats, no adult member is dispensable. These groups are products of the coalition game, and the dynamics of that game create egalitarian pressures.

Equality survives in those contemporary groups whose societies are small and whose relations with neighboring bands are often tense. Our

[35] Here my views are close to those of Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1991).

[36] See Christoph Boehm, *Hierarchy in the Forest* (Cambridge, MA: Harvard University Press, 1999); Richard Lee, *The !Kung San* (Cambridge, UK: Cambridge University Press, 1979); Raymond Firth *We, The Tikopia* (Boston: Beacon, 1961), Marjorie Shostak, *Nisa* (Cambridge, MA: Harvard University Press, 1981).

ancestors lived like that until roughly ten to fifteen thousand years ago. Consequently, more than three-quarters of the period through which the ethical project has evolved was spent in social circumstances now quite rare. Small societies reasonably fear the interference and predations of neighbors. Social cohesion is vital, and no adult can be marginalized in normative discussion. As the coalition game (§9) already revealed, the hominid bands out of which early human societies grew resulted from the partitioning of the physical environment through coalition building. The stability of the partition depends on the approximate balance among neighboring groups, and, where the groups are small, the contribution of every member is necessary. Discussions that involve all adults, that aim to answer to the needs of all adults, and that blur distinctions of rank and ability were crucial to roughly the first forty thousand years of the ethical project.[37]

Those discussions would have issued in agreed-upon rules for life together—but not merely on that. Ethical codes are multidimensional: besides explicit rules, they involve categories for classifying conduct, stories that describe exemplary actions (both commended and frowned upon), patterns of socialization, and habitual forms of behavior. At the earliest stages, we should think of all these elements as accepted by all members of the group. Around the campfires, they reached agreement on precepts, on stories of model behavior, on ways of training the young, on practices of punishment, on sanctioned habits, perhaps occasionally on changes in the concepts hitherto employed. This form of socially embedded normative guidance set the stage for the evolution of the ethical project.

Ethical codes can pronounce on their own amendment, firmly disallowing any possibilities of change or welcoming revisionary discussion. Perhaps at early stages, there was a common insistence on clear rules, to be followed obediently and never to be modified. The difficulties of

[37] My estimates here are speculative. I suppose that the ethical project began with the acquisition of full language, at the latest fifty thousand years ago, and that human societies were small until, at the earliest, fifteen thousand years ago. I conclude that the social egalitarianism observed in contemporary hunter-gatherers, and the kinds of social discussions in which they engage, was central to the ethical project for at least the first thirty-five thousand years.

earlier hominid/human social life were surely sufficiently extensive that initial proposals were incompletely successful, and the social groups that went furthest in resolving their altruism failures almost certainly did so by permitting attempts to adjust what had already been achieved. The codes thus devised and amended are *social* products: they represent a *joint* reaction to the altruism failures previously afflicting the group and they aim to diminish the frequency of similar failures in the future. They presuppose the individual capacity for normative guidance, but how the members are to be guided is a matter for all to decide. The initial function is to reduce the incidence of altruism failures, and codes are fashioned by social apprehension of the ways in which cooperation has broken down.

Does this overemphasize the social character of the ethical project? According to an alternative—"biological"—hypothesis about the origins of ethics, not only did our early human ancestors acquire a disposition to respond to orders—eventually a disposition to command from within—but also the content of the commands given, rather than being fixed through social discussion, embodied shared biases toward particular kinds of rules. Instead of a capacity for normative guidance to be steered in various directions, depending on the ways in which altruism failures are seen as arising (and probably reflecting the actual history of failures of a particular group), the rival conjecture views individuals as evolutionarily biased toward specific modes of self-command.[38]

The biological hypothesis envisions psychological changes. People acquire dispositions to behave in different ways (perhaps sharing more frequently than hitherto), and concomitant capacities to feel particular emotions or to render particular kinds of judgments (negatively directed toward those who do not share). They are furnished with a *moral sense* that redirects some of their own conduct and is expressed in reactions to the actions of others (and sometimes to their own prior behavior). *But the acquisition of this sense would not yet give rise to the ethical project.* Armed with it, members of the group act more frequently in accordance with standards we—we who are participants in the ethical project—

[38] The type of view considered here is most clearly expressed by Marc Hauser. See his *Moral Minds* (New York: Harper Collins, 2006).

approve, but they, the original agents, do not yet have these standards or yet see a distinction between the behavior they used to exhibit and that which they now perform. From our perspective they may be more just than their predecessors, or kindlier perhaps, but this is not an assessment they can make.

For them to initiate the ethical project they must come to see certain types of behavior as exemplary or particular rules as commanding their obedience. Could they derive any such recognitional ability from their own dispositions and capacities, or from reflection on what they are moved to do? How would they come to see one desire or action-prompting emotion as different in status from others? They feel many kinds of sentiments (although the emotions available to them depend on the social environments in which they live), but how do they ascertain which ones belong to the "party of humanity"?[39] To identify something as a genuine command, they need to distinguish commands from other pressures, and the most evident possibility is to identify a *source*—a *commander.* Given their environment, the only available source consists of their fellow group members. If there were an explicit practice of discussing and formulating rules for the group, they would be able to draw the critical distinctions. Nothing else in their psychology or in the ambient environment can confer that ability on them. The ethical project can only begin, then, when normative guidance is socially embedded.[40]

Even if there are dispositions to behave in ways we think of as ethically progressive—to refrain from violence, to share more, to comfort the suffering, or whatever—these are merely "nice tendencies," ways of *conforming* to regularities (regularities the ethical project, once it gets going, will approve), but they are not abilities to *obey* rules or precepts. To be the beginnings of the ethical project they must be coupled to a capacity to discern and be governed by rules and commands that receive some sort of authority. The ethical project requires normative

[39] I borrow the phrase from Hume, *Enquiry Concerning the Principles of Morals* (Indianapolis, IN: Hackett, 1986), 77. It serves as a useful reminder of the fact that those who believe in the existence of particular *moral* sentiments—or *moral* judgments—need to explain how agents are able to identify which ones these are.

[40] There are affinities between the line of argument in this paragraph and Wittgenstein's famous private-language argument (*Philosophical Investigations* §§243 ff.).

guidance, and because there are no rival sources of authority to the group (or some subset of it), it demands that normative guidance be socially embedded.

The biological hypothesis needs further refinement if it is to illuminate any aspect of the ethical project. For the novel capacities it posits depend on the social environment.

Consider various forms of the hypothesis. The very strongest would suppose that human beings acquired a tendency to obey particular kinds of rules—or, more properly, to conform to particular kinds of regularities—quite independently of any social backing for those rules. So, for example, with respect to sharing behavior, it might declare that, beyond the limited primate tendencies for sharing, humans acquired a broader disposition compensating for certain kinds of altruism failures. As noted, in this story, normative guidance is not playing any important role; rather, the more extensive human capacities for sharing result from an extra mechanism for psychological altruism. Possibly, our ancestors acquired some such additional mechanism, but no such mechanism could rival the social inculcation of norms in the complex work of enlarging human cooperative tendencies. That is made plain by the prominent part ethical reminders, whether self-given or public exhortations, play in promoting human cooperation—as well as by the controlled experiments on sharing. Effectively, the strong hypothesis must maintain that the large differences between human and nonhuman forms of psychologically altruistic or behaviorally altruistic behavior come about in two distinct ways, some from a strengthened version of the tendencies to altruism already present in other primates and some from human capacities for self-command.

Weaker versions of the hypothesis suppose that evolution under natural selection has equipped people with biases that operate *through* the capacity for normative guidance. Perhaps human beings, placed in any social environment, will develop to feel specific emotions in response to particular types of behavior—positive emotions to sharing (one's own sharing or the sharing actions of others), negative emotions toward failures to share, for example. Social injunctions that direct sharing will thus be more likely to "take" than putative rules prescribing more selfish courses of conduct. At the extreme, it may be supposed that some sets of commands would be impossible for us to follow; they would be analogous to languages we cannot learn.[41]

Experiments in sharing reveal that, in the actual environments in which people grow up, where they acquire from their societies norms prescribing certain types of sharing, laboratory subjects will share with others and will punish those who do not share.[42] Cross-cultural confirmation of the results takes us a little way across the space of potential environments, but it cannot rule out the possibility that common features of contemporary socialization are playing an important causal role. To demonstrate that contrary behavior is impossible for human beings would require showing that *no* environment allows human development to follow a different path. Conclusions of that form are notoriously hard to defend rigorously, because of our massive ignorance of the potential environments.[43] Additionally, we know already that in some environments—unhealthy ones, to be sure—the norms we are supposedly predisposed to follow are violated by human behavior. The ruthlessly self-directed actions of the Ik, the struggles in concentration camps, and the willingness of subjects in psychological experiments to inflict pain on others remind us that, under the right (or, more properly, the wrong) conditions, the supposedly universal effects will not be forthcoming.[44]

[41] Hauser (*Moral Minds*) uses the analogy, and supposes that there is an ethical counterpart to "universal grammar." For reasons given in the text, I am dubious.

[42] The most systematic body of results comes from the work of Fehr and his associates; see the reference in note 19. Hauser lucidly summarizes this.

[43] The problem is exactly analogous to one that bedevils many sociobiological and genetic determinist claims—the difficulty of extrapolating a norm of reaction from a small sample of cases. For diagnosis, see Kitcher, *Vaulting Ambition: Sociobiology and the Quest for Human Nature* (Cambridge, MA: The MIT Press, 1985) and "Battling the Undead" in Rama Singh, Costas Krimbas, Diane Paul, and John Beatty (eds) *Thinking About Evolution: Historical, Philosophical and Political Perspectives* (Cambridge, UK: Cambridge University Press, 2001, 396–414).

[44] See Colin Turnbull, *The Mountain People* (New York: Simon and Schuster, 1972); Primo Levi, *Survival in Auschwitz* (New York: Touchstone Books, 1996); and John Sabini and Maury Silver, *The Moralities of Everyday Life* (Oxford, UK: Oxford University Press, 1982). Turnbull's ethnography is controversial, but unless all his observations are thoroughly false, there would still be grounds for wondering about the hypothesis that our predispositions make contrary norms impossible for us.

Our tendencies to behavior are most likely quite plastic. Given the hypothetical genomic change that underlies the supposedly broadened altruistic tendencies, there would probably be a range of dispositions to action across the (largely uncharted) space of social environments in which people can live. If the conclusions drawn earlier (§11) about the explanation of the behavior of subjects in experiments on sharing are correct, propensities for conduct are likely to depend on the presence of socially embedded normative guidance and the forms that guidance takes. The weaker version of the biological hypothesis is implausible so long as it insists on a specific type of emotional reaction available across all environments and very particular ways in which that emotional reaction is directed independently of the social milieu.

Far more plausible is the idea that, because of our evolved psychology, not all attempts to inculcate norms will do equally well. Perhaps we do have tendencies for emotional responses to types of actions, so that, in the environments that prevail, following one norm might be uncomfortable for us (in the way experimental subjects feel discomfort as they are following the experimenter's order to inflict "pain"), while following another might be accompanied by feelings of ease. To modify the linguistic analogy, given those social environments so far created, some languages might be more difficult to learn—and some sets of commands similarly hard to follow. Human evolutionary history may have bequeathed to us forms of blindness that make reliable compliance with some prescriptions difficult. Without a proof of impossibility, pragmatism counsels societies to work hard at training their members to follow the precepts they deem most important.

Our early human ancestors, equipped with a capacity for normative guidance, were able to explore various possibilities for social exercise of that capacity. Those explorations proceed along two dimensions, one concerned with the ways in which the young are trained in the ethical code, the other focused on the content of the code. Because we know, as yet, so little about any biases with which our evolutionary past might have equipped us, my account will attend to the more visible, social, features of ethical exploration. To proceed in this way is not to conceive of human beings as infinitely plastic, or (to switch images) as blank slates on which societies can write what they please. The history of the ethical project, from the acquisition of normative guidance to the present, is a history of experiments, carried out by social groups who sometimes may have faced difficulties precisely because they rubbed against the grain of human nature in ways of which neither they nor we are aware.

To recapitulate: hominid societies were confronted with recurrent altruism failures, a predicament limiting their size and level of cooperation. Through the acquisition of normative guidance and its social embedding, these failures could be addressed by elaborating ethical codes. The subsequent ethical project is a sequence of ventures in developing such codes, in which—as the next chapter will explain—the dominant mechanism is a cultural analogue of natural selection. It is possible that a small portion of the original altruism failures were corrected by an alternative mechanism, some strengthening of the altruistic tendencies already present among primates (although where we have evidence for any such mechanism, the effects are specific to a range of contexts).[45] It is also possible that human psychological evolution equipped human beings with biases (as yet uncharted) that interfered with or reinforced specific types of ethical codes. Neither possibility undermines the enterprise of trying to understand the main features of the cultural evolutionary process that the acquisition of normative guidance made possible for us.

[45] A prime example is the case of cooperation in child care. See the references in note 1.