Sentimental Rules

*On the Natural Foundations
of Moral Judgment*

SHAUN NICHOLS

2

Sparks of Benevolence

*The Varied Emotional Responses
to Suffering in Others*

Look Rev, I hate to see a man cry. So shove off out the office;
there's a good chap.

—From Monty Python's "Motor Insurance Sketch"

1. INTRODUCTION

In the last chapter, I argued that core moral judgment depends on both
a normative theory and an affective mechanism that is sensitive to suffer-
ing in others. Because core moral judgment emerges so early, and, in
particular, because it can be present in the absence of sophisticated mind-
reading capacities, we need to determine which affective systems are also
available in young children. Thus, in this chapter I want to explore in
detail the nature of the affective response to suffering in others. As it
happens, the human response to suffering in others is complex and inter-
esting. There are different kinds of response, each of which apparently
emerges fairly early in development. Most of this chapter will be devoted
to delineating the emotional responses to suffering and the mechanisms
underlying these responses. But at the end of the chapter, I will explicate
more fully the relation between these emotional responses and core moral
judgment.

What kind of emotional responses to suffering are available? This
issue has been explored most systematically in the literature on prosocial
behavior, that is, behavior that benefits others. The familiar saw is that
"empathy" is the emotional response underlying prosocial behavior. But
research over the last two decades suggests that the unqualified appeal to
empathy blurs importantly different kinds of emotional responses. More-

over, the appeal to empathy might even neglect important categories of
response to harm in others. For it might turn out that some affective
responses to suffering are not plausibly viewed as empathic. Much of the
task of the present chapter will be to sort this out. As we will see, some of
the sorting can be done fairly quickly. One kind of response, "emotional
contagion," is by now familiar in the empathy literature. An apparently
different emotional response, "personal distress" is fairly easy to depict.
A third kind of response, "concern," presents a much more complicated
set of issues. As a result, the bulk of this chapter will be devoted to setting
out the underpinnings of concern.

2. EMOTIONAL CONTAGION AND PERSONAL DISTRESS

As noted above, it is often held that an empathic response provides the
motivation for prosocial behavior.[1] Most generally, empathy is regarded
as a "vicarious sharing of affect" or an emotional response in which the
emotion is "congruent with the other's emotional state or situation" (Eise-
nberg and Strayer 1987, 3, 5). This definition itself encompasses vastly
different ways of attaining the vicarious sharing of affect. For we might
share in another affective response by perspective taking, that is, by imag-
ining oneself to have the other person's mental states.[2] But a different
way that we arrive at the same affect is by emotional contagion, when we
"catch" another's affect. This phenomenon is familiar in everyday life—
it's the reason for laugh tracks in lame sitcoms. More to our purposes,
we can be brought to tears by seeing or hearing another person cry. This
kind of "reactive crying" is present at birth. In a famous study, Simner
presented newborn infants with tapes of a range of auditory stimuli, in-
cluding spontaneous crying by a newborn infant, spontaneous crying by
a five-month-old infant, a computer generated crying sound, and white
noise of equivalent sound intensity (Simner 1971). Simner found that the
newborns cried significantly more in reaction to the tape of the newborn
crying than in the other conditions. These findings have been replicated
by separate laboratories (Sagi and Hoffman 1976; Martin and Clark
1982). The standard interpretation of these results is that the infants are
experiencing emotional contagion (e.g., Thompson 1987). Because the

---

1. Altruistic behavior is one form of prosocial behavior, but prosocial behavior also encompasses
selfishly motivated behavior that helps others. For example, intentionally helping a stranger merely to
impress onlookers counts as prosocial, though presumably not altruistic.

2. I will delay discussion of perspective taking until section 5.

emotion felt is distress, we might regard the infant's response as "contagious distress."

The notion of "personal distress," which has gained prominence largely as a result of work by C. Daniel Batson, is cast as a "self-oriented" feeling that is caused by distress in others. Personal distress is sometimes interpreted as equivalent to contagious distress (e.g., Preston and de Waal 2002). However, it is not obvious that personal distress felt by the observer is homologous to the distress felt by the victim. On Batson's account, personal distress is characterized by "feelings such as upset, alarm, anxiety, and distress" (1991, 117). And when people feel such upset, alarm or anxiety, it is by no means clear that they are feeling anything that mimics what the victim feels. Consider, for instance, the following items, which are supposed to reflect personal distress on a standard questionnaire designed to track individual differences in empathy:

> In emergency situations, I feel apprehensive and ill-at-ease.
>
> I sometimes feel helpless when I am in the middle of a very emotional situation.
>
> Being in a tense emotional situation scares me.
>
> I tend to lose control during emergencies.
>
> When I see someone who badly needs help in an emergency, I go to pieces. (Davis 1980)

At least often in emergency situations, the victim does not go to pieces. In some emergency contexts, badly injured victims are found unconscious and hence not experiencing any distress. But people can still experience upset, alarm and anxiety upon witnessing such victims. Indeed, it seems plausible that people might have these distress reactions on finding an accident victim who is obviously dead. In these cases, then, personal distress cannot be assimilated to any form of emotional contagion. Nancy Eisenberg and Janet Strayer explicitly depict personal distress as a nonempathic response to others' distress: "When perceiving cues related to another's distress, some people may experience an aversive state such as anxiety or worry that is not congruent with the other's state" (Eisenberg and Strayer 1987, 7). On this construal, personal distress does not count as empathy, because it is not a vicarious sharing of homologous affect.

Although contagious distress and personal distress might have important differences, they share important characteristics as well—they are both self-oriented and they are both rather closely tied to situational cues. For our purposes, these shared features will often be more important than the apparent differences between contagious distress and personal distress. Furthermore, although it is evident that infants experience either contagious distress or personal distress, it is not entirely clear which category best characterizes infant responses to distress in others. In light of this, I will use the term "reactive distress" as a broader category that includes both contagious distress and personal distress. As we will see shortly, it is plausible that there is another kind of emotional response available as well—one that is not so situation dependent as reactive distress. The nature of this kind of response emerges from a consideration of work on altruism. As a result, it will be important, and of independent interest, to take a close look at accounts of the underpinnings of altruistic motivation.

3. CORE CASES OF ALTRUISTIC MOTIVATION

Because most of the remainder of this chapter will focus on the underpinnings of altruistic motivation, it will be useful to provide an overview of what's to come. The literature on altruism is simply enormous, and it spans several disciplines including philosophy, social psychology, developmental psychology, and evolutionary biology. Although I will draw on work from all of these areas, my goal is restricted to the project of determining the cognitive and affective mechanisms underlying basic altruistic motivation. Since numerous cognitive mechanisms play an essential role in generating altruistic behavior—for example, perceptual input systems, attentional mechanisms, motor control systems—it will be important to be a bit more explicit about my explanatory goals. I want to sketch an account of altruistic motivation that addresses two different questions. One question asks which mechanism produces the motivational state itself. In keeping with the prevailing views, I will argue that the motivational state is an affective state, produced by an affective system, the "Concern Mechanism." The other question asks which mindreading mechanisms are required to activate the affective mechanism. For the most part, I will defer discussion of the affective component of altruistic motivation until section 9. Until then, the focus will be on the extent to which mindreading is required for altruistic motivation. I will consider in some detail recent proposals about the mindreading mechanisms underlying altruistic motivation. I will argue against the radical view that mindreading capacities are unnecessary for altruistic motivation. Then I will sketch the more prevalent proposal, that altruistic motivation depends on the capacity for perspective taking. I will maintain that none of the arguments for the perspective-taking account is convincing and that there is considerable evidence that altruistic motivation does not depend on such sophisticated mindreading capacities. Rather, I will suggest that altruistic motivation

depends on a Concern Mechanism that requires only minimal mindreading capacities, for example, the capacity to attribute distress to another. This will be rather important for our purposes, because it will mean that by the time children succeed on the moral/conventional tasks, they have the mindreading abilities required for altruistic motivation.

To begin a discussion of altruistic motivation, it will be helpful to set out some core cases of altruistic behavior. In science in general, it is not always clear at the outset what the core cases are, and new evidence and arguments might alter our conception of what should be included as core cases. The situation is no different in studying altruism, and we may want to revise our view about what the core cases are. Philosophical discussions in this area tend to rely on hypothetical cases of altruism. But because the present goal is to give an account of the psychological mechanisms implicated in actual cases of altruistic behavior, it is important to begin with real cases. To his credit, philosopher Lawrence Blum takes this strategy and offers real examples of helping behaviors that he suggests need to be accommodated by an adequate theory of altruism (Blum 1994). Blum's cases all come from young children. For present purposes, it will suffice to recount just a few of the examples:

1. Sarah at twelve months retrieves a cup for a crying friend. (Blum 1994, 186)
2. Michael at fifteen months brings his teddy bear and security blanket to a crying friend. (Blum 1994, 187)
3. A two-year-old accidentally harms his friend (another two-year-old) who begins to cry. The first child looks concerned and offers the other child a toy. (Blum 1994, 187)

The clearest real-life examples of altruistic behavior in adults come from work on helping behavior in social psychology. Perhaps the best known research on adults' helping behavior is the work on the "bystander effect" by Bibb Latané and John Darley (1968). They found that when there are numerous bystanders, subjects are relatively unlikely to offer assistance to those in need. This finding is often used to draw a rather bleak picture of human altruism (e.g., Campbell 1999). However, focusing on these studies obscures the pervasiveness of human altruism. For it turns out that if subjects perceive unambiguously serious distress cues and there are no bystanders, virtually everyone helps. For instance, in one study, Clark and Word (1974) had each subject engage in a distracter task and as the subject left the experiment, he passed a room in which a man (the experimenter's accomplice) made a sharp cry of pain and then feigned unconsciousness apparently as a result of being shocked by an electronic probe. The researchers found that when the accomplice was no longer touching any of the electronic equipment, all of the subjects of-

fered help. And even when the accomplice was still touching electronic equipment (thus presenting potential danger to the helper), over 90 percent of the subjects offered help (Clark and Word 1974, 282). An adequate account of altruistic motivation should explain the underpinnings of these kinds of helping behaviors.

This list of core cases is, admittedly, rather short. It excludes possible cases of altruistic motivation that do not involve helping others in need. Sometimes people are generous to strangers who are not in need, and I do not mean to suggest that such behaviors cannot be altruistically motivated. However, I think that by focusing on a more limited range of cases, we are more likely to make progress on a psychological account of altruistic motivation. The cases of comforting or helping others in distress form a plausible core because such cases emerge so early in children and they appear to be pervasive among adults. Furthermore, although I'm focusing on a very short list of core cases, these cases already present a fairly daunting task. Devising an account of altruistic motivation that would capture both the child cases and the adult cases would be a considerable advance. Of course, it is possible that the examples from children and the examples from adults cannot be captured by a single account. But all else being equal, an account of altruistic motivation that can capture both of these cases would be preferable to an account that captures only one. I will argue that a close look at the role of mindreading in these cases will provide us with a unified account.

4. ALTRUISM WITHOUT MINDREADING?

In recent attempts to characterize the psychological mechanisms underlying altruistic motivation, one central question is the extent to which the capacity for altruism depends on the capacity for mindreading. We will engage this issue in some detail over the next several sections. To begin, I want to consider the radical view that the capacity for altruism is entirely independent of any capacity for mindreading. There are two versions of the view that are discussed in the recent literature. However, as we will see, neither account fits the evidence well.

4.1. Emotional Contagion

Just as empathy is frequently invoked to explain prosocial behavior in general, it is also invoked to explain altruistic motivation. For instance, Goldman writes, "empathy . . . seems to be a prime mechanism that disposes us toward altruistic behavior" (1993, 358). However, we need to

keep clear the distinction between perspective-taking empathy and emotional-contagion empathy. I will postpone discussion of perspective-taking accounts of prosocial motivation until section 5. For now, I want to consider whether emotional contagion might explain altruistic behavior. The standard view, as noted earlier, is that some capacity for emotional contagion is present at birth as evidenced by the fact that infants will cry when they hear the cries of another infant (Simner 1971; Martin and Clark 1982). If emotional contagion appears this early, it is clear that the capacity for emotional contagion does not require the capacity for perspective taking. Indeed, if the capacity for emotional contagion is present at birth, this capacity is presumably completely independent of mindreading capacities. There is some dispute about when mindreading capacities become available, but all sides agree that newborn babies cannot engage in mindreading.

The capacity for emotional contagion suggests a natural and simple account of altruistic motivation. If the distress of another causes oneself to feel distress, this may provide a motivation to relieve the distress of the other—it will thereby relieve one's own distress. This view has a certain elegance, but it is not easy to find a prominent advocate for the view. Although Goldman maintains that altruistic behavior is generated by empathy, Goldman also maintains that emotional contagion is not genuine empathy (1993). Indeed Goldman's simulation account of empathy (described below in section 5) is implausible as an account of emotional contagion (see Nichols et al. 1996), so it is unlikely that Goldman thinks that altruism derives from emotional contagion. Martin Hoffman, one of the most influential figures in empathy research, has been read as proposing something like the simple emotional contagion view in the following passage: "Empathic distress is unpleasant and helping the victim is usually the best way to get rid of the source. One can also accomplish this by directing one's attention elsewhere and avoiding the expressive and situational cues from the victim" (Hoffman 1981, 52, quoted in Batson 1991, 48). It is not clear that Hoffman is really committed to the simple emotional contagion view, but it is instructive to consider the account in any case.

Notice that on the emotional contagion account of altruistic motivation, mindreading is not essential to altruistic motivation. For emotional contagion need not implicate mindreading processes at all. The distress cues are like bad music that you try to turn off. It requires no knowledge of electronics to be motivated to figure out how to stop the offensive stimuli coming from a stereo—one simply experiments with the various knobs and switches. Failing that, one can just leave the room. Similarly, then, one might find the cries of an infant offensive, so one might try to

figure out how to stop the stimuli. To be sure, mindreading can provide useful tools for stopping the unpleasant stimuli. But on this account, mindreading needn't be essential to triggering the motivation to stop the crying.

This story has a prima facie virtue—emotional contagion is thought to be well within the repertoire of young children who provide some of our core cases of altruistic motivation. So, the emotional contagion account provides an extremely simple explanation of altruistic motivation, and it extends to children in an unproblematic way. Hence, it would seem that our problem is solved. Altruistic motivation does not depend on mindreading at all. Rather, it depends on the rather primitive capacity for emotional contagion.

Things are not so simple, however. For consider that, at least in the core cases of altruism from adults, one way to rid oneself of the unpleasant cues is to leave the situation. But this is not what happened in the core cases noted above. Although the subjects could have eliminated contagious distress by fleeing the situation, almost none of them did so (Clark and Word 1974). The fact that adults often help when they could perfectly well escape has now been extensively explored in the work of Batson and his colleagues (Batson et al. 1981; Batson et al. 1983; Batson 1990, 1991). This research provides powerful evidence that some core cases of altruistic motivation cannot be accommodated by the simple emotional contagion account.

Batson has the broader agenda of defending a perspective-taking account of altruism, which we will consider in section 5, but for present purposes, it will suffice to see how his data undermine the emotional contagion account. In classic social psychological fashion, Batson and his colleagues set up a mock shock methodology. Subjects were told that they would be in a study with another person and that one of them would be picked at random to be the worker and the other would be the observer. The worker would perform tasks while being given electric shock at irregular intervals, and the observer would watch the person performing the task under these aversive conditions. Of course, the real subjects always ended up in the observer condition, and the "worker" was really a confederate. The subjects were then told that they would view the "worker" via closed-circuit television (though it was really a videotape). The experiment manipulated the *ease of escape* for the subjects. In the easy-escape condition, subjects read "Although the worker will be completing between two and ten trials, it will be necessary for you to observe only the first two"; in the difficult escape condition, subjects read "The worker will be completing between two and ten trials, all of which you will observe" (Batson 1991, 114). The subjects subsequently viewed the worker endure two trials

(of the ten trials that the worker had agreed to) in which the worker exhibited considerable discomfort. Subjects were given the opportunity to help out the worker by taking over some of her trials. Using this framework, Batson and colleagues also manipulated the degree of "empathy" in the subjects (see section 7 for details). Across a wide range of studies, they found that subjects in low empathy conditions were much less likely to help when escape was easy. By contrast, subjects in the high empathy condition were equally likely to help whether it was easy to escape or not.

For our purposes, the crucial point is the following. On the emotional contagion model, one should only help when it is harder to escape than it is to help. However, the evidence from Batson and his colleagues suggests that there is an important kind of altruistic motivation that cannot be satisfied by escaping the situation. Hence, this kind of motivation cannot be captured by the emotional contagion model (see also Batson et al. 1981; Batson et al. 1983; Miller et al. 1996, Eisenberg and Fabes 1990). More generally, largely as a result of Batson's work, it is now clear that an adequate account of altruistic motivation needs to accommodate the fact that in core cases of altruism, people often prefer to help even when it is easy to escape.

4.2. Sober and Wilson on Altruistic "Sympathy"

In Sober and Wilson's recent book (1998), they propose an alternative path to altruism that does not rely on mindreading or emotional contagion, but rather on a certain kind of sympathy. They suggest that both sympathy and empathy may motivate altruistic behavior (e.g., 1998, 232). They then try to distinguish sympathy from empathy in two ways.

First, Sober and Wilson maintain that there is a crucial difference between empathy and sympathy because in sympathy,

> your heart can go out to someone without your experiencing anything like a similar emotion. This is clearest when people react to the situations of individuals who are not experiencing emotions at all. Suppose Walter discovers that Wendy is being deceived by her sexually promiscuous husband. Walter may sympathize with Wendy, but this is not because Wendy feels hurt and betrayed. Wendy feels nothing of the kind, because she is not aware of her husband's behavior. It might be replied that Walter's sympathy is based on his imaginative rehearsal of how Wendy would feel if she were to discover her husband's infidelity. Perhaps so—but the fact remains that Walter and Wendy do not feel the same (or similar) emotions. Walter sympathizes; he does not empathize. (1998, 234–35)

But this example does not really distinguish sympathy from empathy. As Sober and Wilson seem to anticipate, a sophisticated empathy account can easily accommodate their case by claiming that we use our imagination to empathize with what Wendy would feel if she were to discover the infidelity. Hence, as far as this example is concerned, "sympathy" might merely be a special form of empathy.

The second, and more important, feature of their account is their claim that "sympathy" does not require mindreading. Sober and Wilson maintain that empathy requires that one be a psychologist, but that sympathy does not: "Empathy entails a belief about the emotions experienced by another person. Empathic individuals are 'psychologists' . . . ; they have beliefs about the mental states of others. Sympathy does not require this. You can sympathize with someone just by being moved by their objective situation; you need not consider their subjective state. Sympathetic individuals have minds, of course; but it is not part of our definition that sympathetic individuals must be psychologists" (1998, 236). Thus, Sober and Wilson apparently maintain that "sympathy" does not require any capacity for mindreading.

Of course, Sober and Wilson are welcome to define a notion of "sympathy" on which mindreading is not required for sympathy. However, they provide no evidence that this kind of sympathy exists. If we rely on traditional signs of sympathy, the evidence suggests that children only begin to exhibit the characteristic signs of sympathy after the first birthday (see section 9) and at this age, children probably have some rudimentary mindreading skills (see, e.g., Gergely et al. 1995; Woodward 1998). So, it may well turn out that the capacity for sympathy exists only in creatures that have mindreading capacities and that the capacity for sympathy depends crucially on the capacity for mindreading. Furthermore, even if Sober and Wilson's "sympathy" does exist, they provide no reason to think that it explains anything like the core cases of altruism with which we began. Again, as we will see, children only begin exhibiting comforting behaviors after the first birthday, by which time they probably have some rudimentary mindreading skills. So, if we take Sober and Wilson's suggestion as an empirical claim about the cognitive underpinnings of core cases of altruistic motivation, it is utterly unsupported.

In sum, then, neither emotional contagion nor Sober and Wilson's sympathy provides a promising explanation of altruistic motivation. It is particularly clear that neither proposal offers a unified account of the core cases of altruistic motivation with which we began. Hence, if we are to have a model of altruistic motivation that can accommodate our core cases, it cannot be one of these models that rejects outright the role of mindreading.

5. PERSPECTIVE-TAKING ACCOUNTS OF ALTRUISTIC MOTIVATION

In the Piaget-Kohlberg tradition, the capacity for perspective taking is thought to be essential to a wide range of moral capacities, including altruistic behavior. Unlike the no-mindreading accounts of altruistic motivation, there is no shortage of advocates for the perspective-taking account of mindreading and altruism. In the recent literature, the most prevalent account of mindreading and altruism continues to be that altruistic motivation depends on perspective taking. This view is suggested by several figures including Batson (1991), Blum (1994), Darwall (1998b), and Goldman (1993).

Goldman (1992; 1993) is by far the most explicit about the cognitive architecture underlying perspective taking, so his work provides a useful starting point. As we've seen, Goldman maintains that empathy is central to altruism, and he maintains that genuine cases of empathy depend on perspective taking. His account of perspective taking draws on his earlier work on the off-line simulation account of folk psychology (Goldman 1989; see also Gordon 1986). Goldman maintains that the process of perspective taking is subserved by off-line simulation in the following way: "Paradigm cases of empathy . . . consist first of taking the perspective of another person, that is, imaginatively assuming one or more of the other person's mental states. . . . The initial 'pretend' states are then operated upon (automatically) by psychological processes, which generate further states that (in favorable cases) are similar to, or homologous to, the target person's states. In central cases of empathy the output states are affective or emotional states" (1993, 351). Now, if we try to incorporate this account of empathy into an account of altruistic motivation, we get the following account of the processes underlying altruistic motivation when the agent sees another in distress. First, the agent determines the beliefs and desires of the person in distress. Then the agent pretends to have those beliefs and desires. These pretend-states are then operated on automatically, leading to affective states that are similar to the target's state, namely distress. These unpleasant affective states then motivate the agent to eliminate the problem at its source—the other person's distress.

Batson's picture is less architecturally explicit, but is still clearly dependent on perspective taking. Batson claims that altruistic motivation derives from "empathy" (1991, 83), and as Batson defines it, empathy requires perspective taking. He writes, "Perception of the other as in need and perspective taking are both necessary for empathy to occur at all" (1991, 85). The empathic response to perceived need "is a result of the perceiver adopting the perspective of the person in need" (1991, 83) and this involves "imagining how that person is affected by his or her situation" (1991, 83).

Blum's (1994) view is somewhat more difficult to interpret. He maintains that altruistic behavior, or "responsiveness" requires "that the child understand the other child's state" (1994 , 197). He rejects the idea that this understanding is limited to cases in which the subject infers "the other's state of mind from a feeling the subject herself has, or has had, in similar circumstances" (1994, 192). Blum rejects this account because it is too "egocentered" (1994, 193), and he argues that this cannot be the sole cognitive process because "such inference would not account for understanding states of mind different from those one is experiencing or has experienced oneself" (1994, 192). Rather, Blum maintains that "understanding others means understanding them precisely *as other* than oneself—as having feelings and thoughts that might be different from what oneself would feel in the same situation" (1994, 193; emphasis in original). So Blum apparently maintains that altruistic motivation depends on the understanding of others as potentially having different beliefs, desires, and emotions. But he does not offer an explicit explanation about how this understanding is achieved.

Although these accounts have important differences, they all share an assumption that altruistic motivation depends on some fairly sophisticated mindreading capacities. First, on Blum's account, and possibly Batson and Goldman's as well, the subject must be able to recognize that the other person might have different mental states than the subject herself would have in a similar situation. Second, for Goldman and Batson, perspective taking requires using the imagination to figure out someone else's mental states. As a result, in sharp contrast to the emotional contagion account, the perspective-taking accounts of altruistic motivation invoke complex mindreading capacities.

6. A MINIMAL MINDREADING ACCOUNT OF ALTRUISTIC MOTIVATION

The accounts of altruistic motivation that make no appeal to mindreading have difficulty accommodating the psychological evidence and capturing the core cases of altruistic motivation. However, I think that we can accommodate the data with a much more austere proposal about the role of mindreading than the perspective-taking accounts. I want to sketch an account of altruistic motivation that draws on as little mindreading as

necessary to accommodate the core cases of altruism. Then in the subsequent sections, we will consider the relative merits of the minimalist account and the perspective-taking account.

The crucial finding on altruistic motivation from social psychology is the fact that people often help even when it would be easy to escape (e.g., Batson 1991). If the motivation is caused strictly by an aversive response to immediate situational cues, as proposed by the simple emotional contagion model, then escape is a good alternative. For one can simply remove oneself from the source of discomfort. However, escape is not an adequate alternative if the motivation comes from an enduring internal cause. As a result, a natural first move is to suppose that subjects elect to help rather than escape because some aspect of the situation is preserved in an enduring mental representation, and this mental representation produces the motivation. One could conceivably try to use this move to extend the emotional contagion account. An emotional contagion theorist might continue to deny any role for mindreading and maintain that altruistic motivation comes from an enduring representation of the behavioral, acoustic, or physiognomic cues that cause contagious distress. On this modified emotional contagion account, the reason subjects do not escape in the experiments is that the motivation comes not simply from the immediate situational cues, but also from the enduring representation of those cues. So, on this story, the subjects help because the emotional contagion can only be alleviated by eliminating the aversive cues. However, even this extended emotional contagion account is still inadequate. The problem is that superficial cues can produce emotional contagion, and if one knows that the cues leading to emotional contagion are merely superficial, this typically does not prevent one from experiencing emotional contagion, but it does undermine altruistic motivation.[3] In the present context, the best way to see the problem is by considering what the account predicts about behavior in Batson-style scenarios with superficial distress cues. The extended emotional contagion account predicts that in these situations, subjects will be motivated to eliminate superficial distress cues rather than escape, and, although the relevant experiments have not been conducted, this prediction seems most implausible. For instance, if a subject found herself in an empty classroom with a projector

showing a computer-generated hologram of a baby crying convulsively, this stimulus would likely produce a negative affective response, and presumably the subject would have enduring representations of the cues that lead to this negative response. But in this case, so long as the subject realizes that the stimulus is a hologram and not a real crying baby, her aversive response will likely be relieved about equally well whether she turns off the projector or leaves the room. As a result, the extended emotional contagion account does not accurately predict when escape will be an adequate solution for the subject.

Rather than opt for this implausible attempt to rescue the emotional contagion view, I think that we need to appeal to some capacity for mindreading to obtain an adequate account of altruistic motivation. A rough first proposal here is that altruistic motivation depends, not on a representation of superficial cues, but on a representation of the target's pain (or some other negative affective or hedonic mental state). Appealing to these kinds of representations will provide at least a partial explanation for why subjects help rather than escape in Batson-style scenarios. If altruistic motivation is triggered by a representation that the target is in pain, escape is not an effective solution to the motivational problem because merely escaping the perceptual cues of pain will not eliminate the consequences of the enduring representation that another is in pain. Thus, this account provides some explanation for why escaping the situation is not an adequate solution. Further, the account explains why the extended emotional contagion account is inadequate—if, in a Batson-scenario, you know that the aversive cues are merely superficial, then you do not have a representation that the target is in pain, so escape is an adequate solution. This account also fits well with Batson's finding that the motivation to help is relieved when the subject comes to think that the target's pain has been alleviated, regardless of whether the target's pain was alleviated by the subject or someone else (Batson 1991).

I suggest, then, that altruistic motivation depends on the minimal mindreading capacity to attribute negative affective or hedonic states to others. On this view, a person can have the capacity for altruistic motivation even if the person does not have or does not exploit the capacity for imagining himself in the other's place and having different beliefs, desires or emotions than he himself would have in that situation. However, a person cannot have the capacity for altruistic motivation without some capacity to attribute negative affective or hedonic states to another. For the remainder of the chapter, I will focus on distress as the exemplar mental state, but this is merely for ease of exposition. I do not mean to exclude the possibility that representations of other negative affective and hedonic states (e.g., grief, fear, sorrow) will produce altruistic motivation.

3. The notion of emotional contagion is defined as involving a "vicarious sharing of affect." As a result, the definition technically precludes merely superficial cues from producing emotional contagion, because one cannot share affect with something that does not have that affect. But obviously the mechanisms underlying emotional contagion can be activated by merely superficial cues, because one can synthetically produce many of the cues that lead to "real" emotional contagion. Since the issue at hand is what mechanisms are in play, it is appropriate to use a broader category of response that also includes instances of "ersatz" emotional contagion in which the affect is produced by merely superficial cues.

Appealing to the capacity to attribute distress helps explain why subjects are motivated to help even when they could more easily escape. Thus, the account seems, at least at this point, to accommodate the important cases promoted by social psychologists. However, I have thus far neglected to consider whether the account fits with the other class of core cases—comforting behaviors in young children. Is there reason to think that young children attribute distress? And are such attributions plausibly connected with their comforting behaviors? The answer to both questions is "Yes." Henry Wellman and colleagues have explored emotion and pain attribution in the spontaneous speech of young children, using transcripts of children's speech from the CHILDES (Children's Language Data Exchange System) database (MacWhinney and Snow 1990). Though this database was initially established to study children's language it has been an extremely valuable resource for studying the young child's understanding of the mind (see especially Bartsch and Wellman 1995). Wellman and colleagues examined the spontaneous speech of five children, focusing on the transcripts collected for each child from the age of two until the age of five. The researchers found that already at the age of two, the children frequently make attributions of pain, usually using the word "hurt" (Wellman et al. 1995, 130). Furthermore, in the cases analyzed by Bartsch and Wellman (1995), there are transcripts available for four children before the age of two, and in each of these cases, the child is attributing pain well before the second birthday (Sachs 1983; Bloom 1970; Bloom 1973). So pain attribution apparently emerges very early indeed.

Not only do young children make pain attributions, but in the work on comforting behavior, we find that young children respond to a variety of distress cues, and they direct their comforting behavior in ways that are appropriate to the target's distress. As we saw in the examples from Blum (1994), children exhibit comforting behavior in response to another's crying. In experimental studies on one-year-olds, crying also elicited comforting behaviors; so did coughing and gagging (Zahn-Waxler and Radke-Yarrow 1982, 116); and Zahn-Waxler and colleagues (1992a) found that children exhibited comforting behaviors in response to the target bumping her head, saying "ow" and rubbing the injured part. Furthermore, in these studies, the children often comfort the target in appropriate ways. Zahn-Waxler and Radke-Yarrow (1982) conducted a longitudinal study in which a group of fifteen-month-olds and twenty-month-olds were each studied for nine months. The researchers report that during this period, every single child in these groups exhibited an instance of "prosocial actions that focus on the specific distress cue" (124). For example, they describe one instance in which the mother of a nineteen-month-old child hurts her foot and the child witnesses the event. The child exhibited con-

cern, ran over, said "hurt foot" and rubbed the mother's hurt foot (124). In addition to showing that young children direct their comforting behaviors in appropriate ways, this example also indicates that young children actually make pain attributions in conjunction with their comforting behavior, and they seem to recognize what the target is distressed about.[4] Thus, there is good reason to think that the minimal mindreading account I've proposed to explain the core cases of altruistic motivation in adults can also be extended to explain the comforting behaviors of young children.

As noted in section 3, this account is not intended to capture all instances of what we would consider altruistic motivation. We can be motivated to be altruistic to someone without attributing any negative affective states to them. For a dramatic example, we might be motivated to prevent the painless death of a peacefully sleeping stranger. However, one of the aims here is to develop an account of altruistic motivation that does not exceed the cognitive abilities of young children who exhibit comforting behavior. And although children display comforting behavior before the age of two, they do not have an understanding of death until much later (see, e.g., Carey 1985). So if we try to develop an account of altruistic motivation that will capture cases like preventing painless death, the account might no longer be able to accommodate young children. As a result, I think that a promising initial strategy in developing a cognitive account of altruistic motivation is to focus on cases of altruistic motivation that are clearly within the repertoire of young children. This will leave

4. One interesting question for future research concerns the extent to which altruistic motivation depends on the child's appreciation that distress is an intentional mental state, a state that is (or can be) directed towards some object. For example, is the altruistic motivation mechanism (the "Concern Mechanism" to be discussed below) activated by the attribution that Mommy is "sad that she hit her foot"? The evidence from Radke-Yarrow and Zahn-Waxler (1982) suggests that children often do appreciate what the target's distress is about, and I mean for the minimalist account to be consistent with the possibility that the motivational system can be activated by attributions of distress as an intentional state. However, it is possible that one might try to develop an even more minimalist account on which altruistic motivation is activated by a general attribution that the target is in distress, with no specification of what the distress is about. (I am indebted to Paul Harris for raising this issue.)

Much of the available evidence seems to be compatible with both of these minimalist alternatives. Consider, for instance, the case from Blum in which the child retrieves a cup for a crying child. One possibility is that the child's belief that the target is "sad that she lost her cup" activates the altruistic motivation system which then produces the motivation to relieve the target's lost-cup distress. An ultra-minimalist might maintain rather that the child's motivation comes from the general attribution that the target is in distress. That is, the attribution that the target is sad (simpliciter) activates the motivation system which produces the motivation to relieve the target's sadness. The child then uses other resources to determine a course of action for relieving the target's sadness, and these resources might include the fuller intentional attribution of the target's specific distress states. It is an open empirical matter which of these stories is right about the requisite mindreading underlying altruistic motivation, and of course it is possible that the altruistic motivation system can process both kinds of distress attribution.

open a number of interesting issues about the relation between "early" altruistic motivation and "mature" altruistic motivation. One possibility is that mature altruistic motivation develops out of the core system that I am attempting to sketch in this chapter. Another possibility is that there are independent systems subserving what we commonly group together as mature altruistic motivation, and the early emerging core system is just one of these independent systems. I will not try to address those issues here. There is, however, another fundamental way in which this account is only a partial account of altruistic motivation. Like the perspective-taking account, the minimal mindreading account does not yet explain the process that goes from mindreading to motivation. As will be discussed below (section 9), on both the perspective-taking account and the minimalist account, a natural assumption is that the representations generated by mindreading produce an affective response that motivates the agent to behavior altruistically. But first, we need to consider the relative merits of the minimal mindreading account and the perspective-taking account.

7. ARGUMENTS FOR PERSPECTIVE TAKING: BATSON'S EVIDENCE

Now that the two proposals are on the table, we can consider the arguments for each account. Although it is widely thought that altruistic motivation depends on perspective taking, it is not easy to find an argument for the view in the recent literature. The only systematic argument comes from Batson's data. Batson used various methods to manipulate the "empathy" of subjects, creating conditions in which subjects would have either high empathy or low empathy. Batson is less architecturally explicit than one would like. But according to Batson, his evidence indicates that perspective taking is required for altruistic motivation because in the experiments high empathy subjects were much more likely than low empathy subjects to help in easy-escape conditions (e.g., Batson 1991, 87; see also Darwall 1998b, 273). Batson's data do, I think, provide an important source of evidence against emotional contagion accounts, but they fall far short of establishing that perspective taking is required for altruistic motivation.

To begin, it is important to note that Batson's experiments cannot be decisive evidence for the perspective-taking account. For the evidence does not show that altruistic motivation is absent among those with low empathy. A substantial minority of subjects in the low empathy conditions do help—averaging across studies, nearly a third of the low empathy

subjects helped (Batson 1991, chapter 8). And it is possible that most of the other low empathy subjects had some altruistic motivation, but not enough to outweigh the competing motivation to avoid the pain of electric shock. Submitting to painful electric shock to relieve a stranger is a rather high cost action, and it seems likely that if the altruistic option were low cost (e.g., returning an elderly person's books to the campus library), then the difference between high empathy and low empathy subjects might largely disappear.

Although Batson's evidence hardly counts as a decisive argument for the perspective-taking account, it does seem that the perspective-taking account provides a natural explanation for why high empathy would lead to higher altruistic motivation. For if altruistic motivation depends on taking the perspective of others, then increased perspective taking might increase the motivation. However, I think that the minimalist account can provide equally good explanations for Batson's findings. To see why, we need to consider in a bit more detail Batson's two central empathy manipulations: the perspective-taking manipulation (Batson 1991, 120) and the similarity manipulation (Batson 1991, 114). In the perspective-taking manipulation, subjects watched a videotape of a student with broken legs. The subjects were either told to "attend carefully to the information presented on the tape" or to "imagine how the person interviewed felt about what happened." Subjects who were told to imagine the other's feelings were more likely than subjects in the other group to help in the easy-escape condition. Although the perspective-taking account can explain these results, the minimalist account can explain the results equally well. For in the high perspective-taking conditions, subjects are more likely to focus on the other's distress, and they are more likely to develop elaborate representations of the other's distress. Thus, on the minimalist account, it is hardly surprising that the perspective-taking manipulation facilitates altruistic motivation, because perspective taking implicates representations of the other's distress. In principle, it will be hard to undermine a minimalist account using this kind of manipulation because if you increase a subject's perspective taking of a distressed target, you will also typically increase the subject's representations of the target's distress.

In Batson's other important "empathy" manipulation, subjects were shown a questionnaire purportedly filled out by the person who would later need help. One group of subjects saw questionnaires that expressed similar views to those expressed on the subject's own questionnaires. The other group saw questionnaires that expressed dissimilar views. Batson and colleagues found that subjects in the high-similarity group were more likely than subjects in the low-similarity group to help in the easy-escape condition. Batson notes that previous research by Stotland (1969) and

Krebs (1975) shows that subjects in high-similarity conditions display increased empathy. But there is a crucial hedge on "empathy" here. What Stotland (1969) and Krebs (1975) found was that subjects in high-similarity conditions showed heightened physiological response and expressed more concern for the other person. The level of perspective taking in these tasks was not measured. Nor do the researchers suggest that perspective taking is the crucial mechanism underlying the response of subjects in high-similarity conditions. There is, in fact, a large literature in social psychology suggesting that subjects are more attracted to people they think have similar attitudes (e.g., Newcombe 1961; Byrne 1971), and even that people are repulsed by those that they think have different attitudes (Rosenbaum 1986). In light of this, it is hard to see how Batson's similarity manipulation could support the perspective-taking account. What his findings do show is that we are more likely to help people who we think have similar attitudes (for a disturbing variation on this, see Tajfel 1981). Coupled with the data on similarity and attraction, we might conclude from this that we are more prone to help people that we like. But this is quite irrelevant to whether altruistic motivation requires perspective taking.

8. DEVELOPMENTAL EVIDENCE AND PERSPECTIVE TAKING

Thus far, we have no reason to think that altruistic motivation depends on the kind of sophisticated mindreading suggested by perspective-taking accounts. In this section, I will argue that the empirical evidence actually weighs against the perspective-taking account. As we saw in the first chapter, in trying to determine the core architecture underlying a capacity, cognitive scientists pay close attention to evidence from development and evidence from psychopathologies. I will argue that evidence from development indicates that altruistic motivation is independent of sophisticated mindreading abilities like perspective taking. In section 10, I will take up evidence from psychopathologies and argue for a similar conclusion.

The discussion of altruism began with Blum's cases of altruism in young children. Nor are his examples atypical. Blum draws some of his examples from a large body of literature in developmental psychology. This research claims that we start seeing the kind of behavior exemplified in Blum's cases early in the second year. Radke-Yarrow and colleagues (1983) found that at ten to twelve months, children did not respond like the kids in Blum's examples, but "Over the next six to eight months the

behavior changed. General agitation began to wane, concerned attention remained prominent, and positive initiations to others in distress began to appear" (Radke-Yarrow, Zahn-Wexler, and Chapman 1983, 481). And, as noted earlier, in Zahn-Waxler and Radke-Yarrow's (1982) study, they found that every single one of their young subjects performed a prosocial act directed at a specific distress cue.

Despite this impressive capacity for altruistic motivation, children under the age of two have severely limited mindreading abilities. Of particular significance, young children have severe deficiencies in their capacity to take the perspective of others. As noted earlier, children under the age of four fail the standard false belief task and similar tasks (see also Wellman 1990; Bartsch and Wellman 1995). Furthermore, although children begin to pretend by around eighteen months, they seem unable to use the imagination to understand other minds until much later (see, e.g., Nichols and Stich 2000, 2003). Thus, since toddlers provide core cases of altruistic motivation and they lack the requisite perspective taking capacities, this provides a serious prima facie argument against the perspective-taking accounts.[5]

In fact, young children's comforting behaviors offer a striking picture of both altruistic motivation and limited perspective taking. The comforting behaviors of young children tend to be "egocentric." Hoffman notes that young children's helping behaviors "consist chiefly of giving the other person what they themselves find most comforting" (1982, 287). For instance, young children will offer their own blanket to a person in distress. Hoffman offers an example of a thirteen-month-old who "responded with a distressed look to an adult who looked sad and then offered the adult his beloved doll" (1982, 287; see also Zahn-Waxler and Radke-Yarrow 1982; Dunn 1988, 97). Thus, toddlers' comforting behavior seems to be simultaneously altruistic in motivation and egocentric in perspective.

Although much early altruistic behavior is guided by "egocentric" considerations, this is perfectly compatible with the minimalist account. A common interpretation of the fact that toddlers offer their own comfort objects is that it shows that children do not really understand that it is the other person who is in distress. For instance, Hoffman (1982) claims that the fact that children tend to give their own comfort objects to help others indicates that "Children cannot yet fully distinguish between their

5. Of course, one might deny that toddler comforting behaviors count as core cases of altruism. Rather, one might claim that such cases should be construed as ersatz altruism. However, one would need an argument for excluding these cases. For if we focus on the underlying motivation, the evidence suggests that altruistic concern in toddlers is continuous with altruistic concern in later childhood and adulthood (e.g., Zahn-Waxler et al., 1992a; Eisenberg and Fabes, 1990; Eisenberg et al., 1989).

own and the other person's inner states . . . and are apt to confuse them with their own" (1982, 287). However, the examples of "egocentric" comforting responses provide no reason to think that the child fails to distinguish her own states from the states of others. On the contrary, these responses provide evidence that the child recognizes that the other is in distress. After all, the child is offering the comfort object to the *other* person. Further, the fact that the child offers a comfort object suggests that the child does understand that *distress* is involved. Children do not try to relieve the other's distress by completely bizarre behavior like pretending that a banana is a telephone. And there's no reason to think that before eighteen months, the child experimented with various means of eliminating crying in others (as one might experiment with an unfamiliar piece of electronics). However, the young child has limited mindreading resources at hand and thus relies on egocentric mindreading strategies. As a result, the child's knowledge of how his distress is relieved guides his thinking about how to relieve the other person's distress. Thus, the toddler's egocentric comforting cases are not only consistent with the minimalist account, the cases provide further evidence that the child attributes distress to the other person.

Although there is strong evidence against the perspective-taking model, it would be derelict to claim a quick victory for the minimalist account that I have proposed. For there is a less austere alternative that is not excluded by the evidence. By the time toddlers exhibit comforting behaviors, they probably have the capacity to attribute desires that they do not have (see, e.g., Repacholi and Gopnik 1997). So one might maintain that it is this mindreading capacity, the capacity to attribute discrepant desires, that is essential for altruistic motivation. This view has not been elaborated and defended in the literature, but it is possible that the view is close to Blum's (1994) account. Recall that Blum maintains that the understanding of others required for altruistic motivation depends on understanding that others might have thoughts and feelings that are "different from what oneself would feel in the same situation." He rejects more austere accounts as too "egocentered" (193).

Although this moderate "discrepant desire" position does not contravene any of the data, it is unclear why the capacity to attribute discrepant desires should be essential to altruistic motivation. To see this, it is important to distinguish between three different kinds of egocentrism. One kind of egocentrism is just the view that an individual's basic motivations derive solely from that individual's own affective or hedonic states. We might call this view "psychological egoism." Psychological egoism might be wrong, but the issue belongs primarily to the foundations of psychological science, not to moral psychology. On the second kind of egocen-

trism—let's call it "ethical egocentrism"—a person is egocentric if none of the individual's desires are directed at another person's needs, except insofar as the individual thinks that addressing the other person's needs will help him.[6] What is crucial about ethical egocentrism (and what distinguishes it from simple psychological egoism) is that if a person is ethically egocentric, he must go through a process of instrumental reasoning before arriving at a motivation to help another. For he must think that helping another will benefit himself. Both of these kinds of egocentrism need to be distinguished from a third kind of egocentrism—mindreading egocentrism. To say that someone is egocentric in this sense is to claim that the individual either cannot or tends not to grasp that others have different likes and dislikes, different judgments, and different feelings than the individual himself. Notice that ethical egocentrism and mindreading egocentrism make independent claims. A person can perfectly well be ethically egocentric without being an egocentric mindreader. That is, a person might know that others have different interests and beliefs than he does, but at the same time, he might not care in the least about the interests of others, except insofar as he thinks it will affect him. Psychopaths seem to fit this characterization fairly well. Conversely, a person could be an egocentric mindreader without being ethically egocentric. That is, a person might be oblivious to the fact that others have different desires and thoughts than she does, but she might care about trying to help others in need, even if she does not think that doing so will serve her own interests. Of course, if she is an egocentric mindreader, she may not be very effective at helping others, because she will not be sensitive to the variation in desires, feelings, and thoughts that actually exists among those she tries to help. Now, finally, we can get to the point of drawing these distinctions—if someone is an egocentric mindreader, that provides no reason to conclude that she lacks altruistic motivation. The kind of egocentrism that undermines the claim for altruistic motivation is *ethical* egocentrism, not *mindreading* egocentrism. As we've seen, when toddlers offer comfort, they often offer their own comfort objects to others. The fact that these children are using egocentric mindreading strategies does not undermine the claim that these children are altruistically motivated. Even if children turned out to be completely egocentric mindreaders, I see no reason to conclude that their attempts to comfort adults with their dolls and blankets would not be the product of altruistic motivation. Thus, although the discrepant desire view fits with the available evi-

6. Ethical egocentrism is, like psychological egoism, a descriptive claim. Thus, this notion should not be confused with ethical egoism, which is a normative theory according to which the moral worth of a person's action depends only on the consequences of the act for the actor himself.

dence, it is not at all clear why we should prefer this account to the simpler minimalist theory.

9. AFFECT AND ALTRUISTIC MOTIVATION

I have argued that altruistic motivation requires only the minimal mindreading capacity for distress attribution, but I have said nothing about how attributing distress to another leads to altruistic motivation. In keeping with most other accounts, I will assume that altruistic motivation is mediated by an affective response (see, e.g., Eisenberg 1992; Goldman 1993; Hoffman 1991). In this section, I will try to characterize the affective response underlying altruistic motivation.

Before continuing, I should acknowledge that it is possible that affect plays no role in altruistic motivation. Rather, perhaps altruistic motivation follows directly from an attribution of distress. Something like this might be Sober and Wilson's ultimate view (1998, 312ff.). They suggest that evolution built a mechanism for altruistic motivation that does not rely on hedonic or affective states. However, they do not explain how that mechanism might have evolved in the existing motivational systems of our ancestors. Furthermore, there is some evidence suggesting important correlations between affect and altruistic behavior. As I will elaborate shortly, the developmental data suggest a correlation between affective response and helping behavior in children, and the social psychological data suggest a similar correlation in adults. In addition, as we will see in section 10, evidence on psychopathy indicates that psychopaths' lack of helping behavior might be correlated with a deficit to their affective response to others' distress.

If altruistic motivation does depend on affect, what is the character of this affective response? In section 2, I distinguished between personal distress and contagious distress. We have already seen that contagious distress seems an inadequate model for altruistic motivation. Batson also provides intriguing evidence that neither is personal distress the predominant emotion that drives altruistic behavior. There are two manipulations of his basic experimental procedure that provide evidence for this. As detailed in section 4, Batson's basic experimental design has subjects watch the "worker" (actually a confederate) endure some trials of shocks in which the worker exhibits discomfort. In the difficult escape condition, subjects were required to watch the remaining trials; in the easy escape condition, subjects were not required to watch the remaining trials. Subjects were then given the opportunity to help out the worker by taking over some of her trials. In one experiment, after viewing the worker endure the two trials (and before they were given the opportunity to help), subjects were asked to report their feelings in reaction to observing the worker. Subjects' self reports were classed into two groups: (i) those who reported feeling predominantly "personal distress," that is, "self-oriented feelings such as upset, alarm, anxiety, and distress" (117) and (ii) those who reported feeling predominantly "other-oriented feelings" for the victims, like compassion, tenderness, and softheartedness (117).[7] Now for the results: In the difficult escape condition, subjects who reported feeling personally distressed were equally likely to help as subjects who reported feeling compassion. However, when escape was easy, subjects that reported feeling personally distressed were significantly less likely to help than subjects who reported feeling compassion (Batson et al. 1983; Batson 1991, 124).

The other relevant manipulation of Batson's exploits the venerable but invariably surprising misattribution methodology. Batson suggests that for most people, witnessing another in distress likely produces both personal distress and compassion. Hence, Batson reasoned, if subjects could be led to misattribute one of these feelings to a drug, then those subjects would discount this feeling and regard the other feeling as their predominant response. In an experiment designed along these lines, all subjects took a drug called "Millentana" (actually a placebo) before they observed the worker endure any shocks. Half of the subjects were told, "Millentana produces a clear feeling of warmth and sensitivity, a feeling similar to that you might experience while reading a particularly touching novel." Subjects in this condition were expected to regard personal distress as their predominant emotion. The other half of the subjects were told "Millentana produces a clear feeling of uneasiness and discomfort, a feeling similar to that you might experience while reading a particularly distressing novel" (Batson et al. 1981, 298–99). Subjects in this condition were expected to regard compassion as their predominant emotion, because they would misattribute their genuine personal distress to the drug. After observing the worker go through two trials of shocks, subjects were given the opportunity to help. The results came out as predicted. In the easy escape condition, subjects who were led to attribute their personal distress to the drug were more likely to help than subjects who were led to attribute their compassion to the drug (Batson et al. 1981).

These data indicate, then, that the affective response underlying altruistic motivation is not personal distress. However, there remain a couple of importantly different possibilities for the character of the affective re-

7. Batson actually uses the term "empathy" to describe the emotional response of the second group. Obviously, that term only muddies our waters.

sponse. I have suggested that the attribution of another's distress produces an affective response that underlies altruistic motivation. One possibility is that the representation of the other's distress produces a distinctive emotion of sympathy for the other person and this emotion is not homologous to the emotion of the person in need. The sympathy view has some support from an emerging body of research which ties altruistic behavior to a distinctive facial expression (Roberts and Strayer 1996, 456; Eisenberg et al. 1989, 58; Miller et al. 1996, 213). There is also a bit of evidence that sympathy might have distinctive physiological characteristics (Eisenberg and Fabes 1990, 140; Miller et al. 1996). Facial expression and physiological signs are two of the central features that have been used to delineate "basic emotions" (e.g., Ekman 1992). The exciting possibility here is that sympathy is a genuine, distinctive basic emotion with a characteristic facial expression and physiological profile and that this emotion is the motivation behind altruistic behavior. Darwin himself actually made a similar suggestion: "Sympathy with the distresses of others, even with the imaginary distresses of a heroine in a pathetic story, for whom we feel no affection, readily excites tears. . . . Sympathy appears to constitute a separate or distinct emotion" (Darwin [1872] 1965, 215). But Darwin seems to have had a somewhat different notion of sympathy in mind because he thinks that we can sympathize with the happiness of others.

The possibility that altruistic motivation derives from a distinctive basic emotion of sympathy is theoretically appealing, but it has turned out to be difficult to get unequivocal data correlating the postulated features of sympathy with altruistic behavior. There are several different measures, including self-report, facial expressions, and physiological measures. The findings suggest that some of these features are correlated with altruistic behavior some of the time. For example, Eisenberg and Fabes (1990) showed preschoolers a film of children who were injured and in the hospital, and the preschoolers were given the chance to help the hospitalized children by packing crayons for them rather than playing. Although children's self-reports were unrelated to their helping behavior, the physiological measure of sympathy (heart-rate deceleration) was positively correlated with higher levels of helping (Eisenberg and Fabes 1990, 140–41). Further, facial expressions of concerned attention have been significantly correlated with greater helping in boys, but the findings are much weaker for girls (Eisenberg and Fabes 1990, 141). And there is a bit of evidence that there is a correlation between these emotions and helping behavior in Batson-style experiments (Eisenberg et al. 1989).

Notice that the above account suggests that sympathetic motivation for altruism does not count as empathy at all. Rather, on the sympathy

view, altruistic behavior is motivated by a distinctive emotion that is not homologous to the emotion felt by the person in need, or indeed homologous to any other emotion.[8] This would entail that a certain class of empathy-based accounts is thoroughly mistaken. If empathy is a vicarious feeling of the emotion that the target is feeling (caused by perspective taking or emotional contagion), then the empathy account is wrong not just about the mindreading required for altruistic motivation but also about the affect. For on the sympathy account, the emotion driving altruistic behavior does not parallel any other emotion. So, except in the iterative case of empathizing with someone feeling sympathy, empathy will not produce the emotion that generates altruistic behavior.

Although the idea that a distinctive emotion of sympathy underlies altruism is theoretically appealing, there is another possibility. The distress attribution might produce a kind of second order contagious distress in the subject. For example, representing the sorrow of the target might lead one to feel sorrow. This would provide a kind of empathic motivation for helping. And the motivation would be effective even when escape is easy. For the cause of the emotion is still the representation of the other's mental state and as a result, one is motivated not simply to escape the situation because that would not rid one of the representation. As a result, this story would provide an equally effective explanation of the fact that subjects help even in easy escape conditions. And some of the above research on sympathy actually provides support for this alternative story. For instance, Eisenberg and colleagues (1989) found that the strongest predictor of helping in adults was not facial sympathy, but facial sadness (Eisenberg et al. 1989, 61). The available evidence does not really decide between these two accounts of the affect underlying altruistic motivation. Indeed, perhaps both affective mechanisms are operative.[9]

8. As we saw in section 3, Sober and Wilson (1998, 234–5) maintain that sympathy does not require that the sympathizer and the target feel the same emotion simultaneously. But that does not really distinguish sympathy from sophisticated accounts of empathy. The psychological work, however, raises the possibility of a profound distinction. Feelings of sympathy may not parallel any other feeling.

9. As Paul Harris has reminded me, there is a great deal of individual variation among young children in their response to distress in others (e.g., Cummings et al. 1986; Zahn-Waxler et al. 1979). This might be thought to undermine the suggestion that a basic emotion underlies altruistic behavior. However even in emotions that are widely accepted to be basic emotions, one finds considerable individual variation (e.g., Haidt, McCauley, and Rozin 1994 on individual variations in disgust). More importantly, as is often the case in cognitive science, it is difficult to know what to make of the individual variation in behaviors. There is an abundance of factors that seem to contribute to the individual variation in children's responses to another's distress. Some of the variation is attributed to differences in child-rearing practices (Zahn-Waxler et al. 1979); some of it is attributed to other family environmental features (e.g., Klimes-Dougan and Kistner 1990); some of the variation seems to be genetically based (Zahn-Waxler et al. 1992b), and, in older children, some of the variation might be due to differences in perspective-taking abilities (Stewart and Marvin 1984). In light of the complex interaction of these and other

TABLE 2.1 Responses to suffering

	Reactive Distress	Concern
Empathic	Contagious Distress	2nd-order Contagious Distress
Non-empathic	Personal Distress	Sympathy

In section 2, I introduced the term 'reactive distress' to enfold both contagious distress and personal distress. Here I want to make a parallel terminological grouping. I will use the term 'concern' to pick out the class of affective responses that includes both sympathy and second-order contagious distress. For, as we have seen, it is plausible that at least one of these responses explains altruistic motivation, which distinguishes the class crucially from reactive distress. However, it is not yet clear whether altruistic motivation derives from sympathy, second-order contagious distress, or both, so it is useful to lump the responses together under a single heading, 'concern.'

Both contagious distress and second-order contagious distress are empathic. For in both cases, the subject feels an emotion homologous to that of the sufferer. By contrast, neither personal distress nor sympathy is empathic. For in cases of personal distress and sympathy, while the subject's emotion is triggered by suffering (or cues of suffering), the subject's emotion itself is not homologous to that of the sufferer. Hence, these four candidate responses can be slotted neatly into a table (see table 2.1).

It is clear that at least one form of reactive distress is present in earliest infancy. It is also plausible that at least one form of concern is present by the second birthday. In adults, both forms of reactive distress are present; in the case of concern, it is less clear whether both forms are present. But that need not prevent us from trying to characterize more fully the underpinnings of concern.

10. THE CONCERN MECHANISM

We are now in a position to state the proposal about the core architecture a bit more precisely. Altruistic motivation depends on a mechanism that takes as input representations that attribute distress, for example, "John

features, the individual variation seems consistent with the proposal that altruistic motivation depends on a basic affective mechanism.

is experiencing painful shock," and produces as output affect that, inter alia, motivates altruistic behavior. Following the terminology introduced above, I will call this system the Concern Mechanism. In this section, I want to provide a somewhat fuller characterization of the Concern Mechanism, and I will begin by revisiting the perspective-taking account. For there seems to be a double dissociation between the capacity for perspective taking and the capacity for concern.

First, let us return to the developmental evidence. The comforting behaviors of toddlers suggest that the Concern Mechanism is intact and functioning in young children. This is corroborated by a study in which Zahn-Waxler and colleagues traced the development of concern and comforting behaviors in one year old children. They trained mothers to record their child's emotional and behavioral responses to distress in others. Mothers were also trained to simulate various distress situations. Between thirteen and fifteen months, children were reported to respond with "empathic concern" (sad facial expressions or sympathetic remarks) to 9 percent of the natural distress situations and 8 percent of the simulated distress situations. Between eighteen and twenty months, children responded with empathic concern to 10 percent and 23 percent of natural and simulated distress situations. And by twenty-three to twenty-five months, children responded this way to 25 percent and 27 percent of natural and simulated distress situations (Zahn-Waxler et al. 1992a, 131). So it certainly appears that the capacity for concern emerges before the age of two. Furthermore, between eighteen and twenty months, there is a marginally significant correlation between concern and comforting behavior, and by twenty-three to twenty-five months, there is a significant correlation between concern and comforting behavior. The developmental pattern charted by these results suggests, perhaps not surprisingly, that the coordination of the concern response and altruistic behavior is a complicated developmental process. This developmental process no doubt depends on a suite of conditions, environmental and otherwise, that we do not understand. Nonetheless, the broad pattern indicates that the Concern Mechanism is up and running well before the capacity for perspective taking has developed, which suggests that the Concern Mechanism is dissociable from the capacity for perspective taking.

The possibility of a dissociation between the Concern Mechanism and the capacity for perspective taking is further suggested by evidence on children with autism. As discussed in chapter 1, researchers in the mindreading tradition have explored in some detail the capacities of people with autism, and on a wide range of mindreading and perspective-taking tasks, children with autism tend to perform much worse than their

mental age peers (see, e.g., Baron-Cohen 1995; Frith 1989; Dawson and Fernald 1987). Further, one of the central characteristics of autism is lack of imaginative activities and spontaneous pretend play (Wing and Gould 1979). Thus, there is considerable evidence that the capacity for perspective taking is seriously compromised in autism.

Despite their difficulties with perspective taking and imagination, recent studies show that autistic children *are* responsive to distress in others (Bacon et al. 1998; Blair 1999a; Yirmiya et al., 1992). For instance, in one recent experiment, autistic children were shown pictures of threatening faces and distressed faces, and the autistic children showed the normal pattern of heightened physiological response to both sets of stimuli (Blair 1999a). Thus, although autistic children have a deficit in perspective taking, they do respond to the distress of others. In addition, a recent study suggests that autistic individuals engage in comforting behaviors. Sigman and colleagues (1992) explored responses to distress in autistic, Down syndrome, and normally developing children. In one task, the distress was made as salient as possible. The parent was seated next to her child at a small table, and while showing the child how to use a hammer with a pounding toy, the parent pretended to hurt her finger by hitting it with the hammer. The parent then made facial and vocal expressions of distress but did not utter any words (Sigman et al. 1992, 798). Researchers found that autistic children were much less likely than other children to attend to the distress. This fits with a broader pattern of inattentiveness to social cues in autism. For instance, autistic children are much less likely than Down Syndrome children to orient to someone clapping or calling their name (Dawson et al. 1998). Despite the fact that autistic children were less likely to notice or attend to the distress, several autistic children provided comfort to the parent in this experiment. Overall, few children helped, but autistic children helped as often as the children in the other groups.[10]

The fact that autistic children show normal physiological response to distress in others and the finding that autistic children do engage in comforting behaviors suggests that the core architecture for altruistic motivation may be intact in autism. This would pose a serious problem for the perspective-taking account because that account predicts that individuals with serious deficits to imagination and perspective taking would show corollary deficits to altruistic motivation.[11]

So, even though autistic children have a profound deficit in perspective taking, the available evidence indicates no correspondingly serious deficit to the Concern Mechanism. The complementary question is whether there are individuals who show a deficit to the Concern Mechanism but no serious deficit to perspective taking. There's some reason to think that psychopaths fit this description. The standard diagnostic tool used in the United States, the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV*), uses the diagnostic category of Antisocial Personality Disorder, and the *DSM-IV* suggests that psychopathy is the same condition (645). People with Antisocial Personality Disorder "frequently . . . tend to be callous, cynical, and contemptuous of the feelings, rights, and sufferings of others" (647). "Persons with this disorder disregard the wishes, rights, or feelings of others. They are frequently deceitful and manipulative in order to gain personal profit or pleasure (e.g., to obtain money, sex, or power). . . . They may believe that everyone is out to 'help number one' and that one should stop at nothing to avoid being pushed around" (646). A number of researchers characterize psychopathy somewhat differently from the Antisocial Personality Disorder (e.g., Hare 1991), but the alternative diagnostic criteria tend to present a similarly disturbing portrait of psychopaths. For instance, psychopathy is characterized by a lack of remorse and empathy, being deceitful and manipulative, and a tendency to adult antisocial behavior (Hare 1991). These characterizations certainly suggest that psychopaths are significantly less likely than nonpsychopaths to exhibit altruistic behavior. Of course, we have already seen that psychopaths show abnormally low physiological response to suffering in others.[12] Coupled with their apparent lack of altruistic behavior, this suggests that the Concern Mechanism is defective in psychopathy. Nonetheless, evidence indicates that psychopaths are capable of perspective taking, and that they perform as well as normal adults on standard perspective-taking tasks (Blair et al. 1996).

Hence, although the evidence is still preliminary, there seems to be a double dissociation between perspective taking and the Concern Mechanism. Young children and autistic children have immature or impaired perspective-taking abilities, yet young children and perhaps even autistic children have an intact and functioning Concern Mechanism. Psycho-

---

10. Six out of twenty-nine autistic children helped; seven out of thirty mentally retarded children helped; and three out of thirty normally developing children helped (Sigman et al. 1992, 800).

11. Although the evidence on autistic children might pose a serious problem for perspective-taking accounts, it is perfectly compatible with the minimal mindreading account. For as noted in chapter 1,

---

despite their deficits in perspective taking, children with autism are capable of attributing simple negative emotions (e.g., Yirmiya et al. 1992; Baron-Cohen 1995; Tager-Flusberg 1993).

12. It remains to be seen exactly how the affective deficits in psychopathy map onto the different responses to suffering in others that we have been exploring in this chapter. But one story is that both the system for contagious distress and the Concern Mechanism are defective, perhaps because the system for contagious distress is developmentally necessary for acquiring a normal Concern Mechanism.

paths, by contrast, seem to have a normal capacity for perspective taking but a deficit to the Concern Mechanism. The evidence from development and psychopathologies thus counts heavily against the perspective-taking account. It seems that altruistic motivation does not require sophisticated mindreading or perspective-taking abilities. And it doesn't take any imagination to be an altruist.[13]

11. EVOLUTIONARY PRECURSORS

Thus far in this chapter, I have characterized two different classes of responses to suffering in others: reactive distress and concern. Neither reactive distress nor concern requires sophisticated mindreading abilities. Reactive distress can apparently be triggered by rather low-level cues like crying. Concern, I have argued, does require mindreading, but only the minimal mindreading capacity to attribute pain to others. There is reason to think that at least some of these emotional responses are present in our evolutionary forebears.

First, it is important to note that if the basic story about mindreading and the Concern Mechanism is right, this has interesting implications for the possibility of altruism in nonhuman animals. For if human altruism requires so little mindreading, it becomes possible that the mechanisms underlying helping-behavior in some nonhuman animals are analogous to the mechanisms underlying altruistic motivation in humans. Although it is hotly debated at present, some nonhuman animals may well have the mindreading capacity to attribute distress to another. There is some evidence, for instance, that chimpanzees can attribute goals (Premack and Woodruff 1978; Call and Tomasello 1998).[14]

13. The Concern Mechanism has many of the features of modules as set out by Fodor (1983). The evidence on development and psychopathology indicates that the Concern Mechanism has a characteristic ontogeny and a characteristic pattern of breakdown. It is also plausible that the mechanism is fast. It is somewhat more difficult to evaluate whether the Concern Mechanism is "encapsulated" (Fodor 1983, 2000) because the relationship between affective systems and encapsulation is far from clear in the current literature. But the Concern Mechanism plausibly possesses at least one feature of encapsulated systems. A cognitive mechanism is encapsulated if it has little or no access to information outside of its own proprietary database, and one of the central features of an encapsulated system is that such systems resist our preferences (Fodor 2000, 63): You cannot make the Müller-Lyer illusion disappear by wanting it to go away. It is likely that the Concern Mechanism is similarly resistant to our preferences and to the dictates of practical reason. For instance, I might think it best, all things considered, not to feel concern when my daughter gets inoculated because any show of concern on my part might intensify her anxiety about inoculations. Nonetheless, it can be extremely difficult to suppress concern in these circumstances. In this sense at least, the Concern Mechanism resembles encapsulated systems.

14. Apart from its intrinsic interest, the possibility that the psychological underpinnings of altruism might be present in nonhumans is of some importance to an evolutionary approach to altruism. If

Research on nonhuman primates does indicate that at least some nonhuman primates are sensitive to a conspecific's distress signals (e.g., Miller et al. 1963). Preston and de Waal (2002) provide a fascinating review of the experimental literature on the response to another's distress in nonhuman animals. They note that several species have been shown to be upset when witnessing the distress of a conspecific. One particularly striking finding comes from research on rhesus monkeys by Masserman and colleagues (1964). The researchers trained the monkeys to get a food pellet by pulling one chain in response to a red light and a different chain in response to a blue light. After the training session, one of the food-chains was rigged so that it would also administer a sharp shock to a conspecific on the other side of a plexiglas divider. There were fifteen animals in the experiment. Ten of these showed a statistically significant preference for the nonshock chain. Two additional animals stopped pulling either chain for a number of days. The researchers conclude that it seems that the monkeys "will consistently suffer hunger rather than secure food at the expense of electroshock to a conspecific" (Masserman, Wechkin, and Terris 1964, 585; see also Wechkin, Masserman, and Terris 1964).

Thus, there is some evidence showing that some animals are sensitive to the suffering of conspecifics. Nonetheless, it is unclear from the available data which mechanism is operative in the nonhuman primates, whether it is a form of reactive distress or a form of concern. But what is clear and important is that there are powerful evolutionary precursors for responding to suffering in conspecifics.

Although there are evidently evolutionary precursors, I demur from embracing an account of the evolutionary function of these mechanisms. The familiar problem with developing such accounts is a lack of adequate comparative data (see, e.g., Grantham and Nichols 1999; Nichols and Grantham 2000). It is difficult to evaluate, let alone defend, proposals about the evolutionary function of a mechanism in the absence of comparative evidence, including evidence of which animals lack the mechanism. So although we have evidence that many species do respond to distress, it is hard to know the function of the underlying mechanisms

altruistic motivation in humans is an adaptation that depends on sophisticated mindreading abilities like perspective taking, then the altruistic motivation system must have been shaped after the evolution of our sophisticated mindreading abilities. If so, the mechanisms for altruistic motivation presumably emerged relatively recently in evolutionary time because, by most accounts, humans are the only primates with sophisticated mindreading abilities. The Concern Mechanism account of altruistic motivation, on the other hand, need not be committed to the view that altruistic motivation is a recent adaptation because on this view the requisite mindreading mechanisms are minimal and may well have been present in our more distant phylogenetic ancestors.

without knowing which species lack the mechanisms. Experimental questions leap to mind. Do asocial species respond to distress in nonkin? Does the level of response vary with the degree of sociality? Are there gender differences in the response to distress in some species? These questions are much easier to ask than to answer. But without some answers, it is hard to evaluate proposals about evolutionary function in any remotely rigorous fashion. It is for this reason that I have opted to take a descriptive approach to characterizing the mechanisms, rather than an evolutionary psychological approach.

12. SENTIMENTAL RULES AND THE MORAL SENSE

Now that we have a sharper picture of the human responses to suffering, we can provide a more articulate rendering of the Sentimental Rules proposal. According to the Sentimental Rules account, norms that prohibit emotionally upsetting actions receive a special status. Such norms are distinguished from emotionally neutral conventional rules. This chapter makes clear that actions that cause suffering in others are indeed emotionally upsetting in multiple ways. For suffering in others itself is upsetting in multiple ways—suffering or indications of suffering trigger contagious distress, personal distress, and concern. As a result, norms prohibiting harm will, for normal people, be central instances of Sentimental Rules.

With the characterization of the candidate affective mechanisms in hand, it is clear that the Sentimental Rules account differs in important ways from the more traditional sentimentalist view that moral judgment derives from a moral sense. Traditional moral sense theorists, like Hutcheson, promoted the moral sense as the source of distinctive feelings of approval and disapproval which are triggered by the perception of virtue and vice. This moral sense produces a pain of disapproval when we perceive an action that is vicious or morally wrong. Furthermore, when an action prompts the moral sense to deliver the pain of disapproval, we condemn the action as wrong.

None of the affective mechanisms explored in this chapter counts as a moral sense. For none of these mechanisms tracks vice (or virtue). Rather, each of the emotions we have explored—personal distress, contagious distress, and some form of concern—can be activated in the absence of any moral judgment. Eighteen-month-old human infants can probably respond with reactive distress and concern; yet it would be a stretch to say that such infants have the capacity for core moral judgment. More importantly, in adults with a mature capacity for moral judgment,

forms of reactive distress and concern can be activated in the conspicuous absence of any judgment that a transgression has occurred. Indeed, the emotions can be activated in the absence of any judgment that an *action* has occurred. When we come upon accident victims, we exhibit characteristic emotional responses of concern and reactive distress. Yet at least often, this is not accompanied by any judgment of wrongdoing or viciousness.

On the Sentimental Rules account, reactive distress or concern plays a crucial role in leading people to treat harmful transgressions as wrong in a distinctive way. Thus, these relatively simple, primitive emotions supply the sentiment to moral judgment. No further moral feeling is invoked as a necessary part of core moral judgment. Perhaps there is a further special feeling, but it is not posited by the Sentimental Rules account. Rather, on this account, the relatively primitive emotions of reactive distress and concern lead us to treat harm norms as distinctive. But again, these emotions can be activated in the absence of a judgment of vice. So none of them can be identified with a moral sense.

13. CONCLUSION

The human response to suffering is impressively multifaceted. Contagious distress, personal distress, and some form of concern each constitute distinctive emotional responses elicited by suffering in others. Some of these responses emerge early ontogenetically, perhaps even by the time of birth. But all of them are plausibly present in the young child, including the elements of altruistic motivation. The evidence suggests that basic altruistic motivation requires only a minimal capacity for mindreading, the capacity to attribute negative affective or hedonic mental states like distress. These attributions, I have suggested, produce altruistic motivation by activating an affective system, the Concern Mechanism. Of course, the account of altruistic motivation I have offered in this chapter is hardly a full account of the psychological mechanisms implicated in mature altruistic behavior, for the altruistic capacities of adult humans far outstrip those provided by the primitive mindreading and Concern mechanisms. Nonetheless, the empirical work suggests that the Concern Mechanism and a minimal capacity for mindreading form the core of our capacity for altruism. By the age of two years, this capacity is in place in normal humans, along with the capacity for reactive distress, which is in place much earlier.

The emotions set out in this chapter allow us to fill out the Sentimental Rules account of core moral judgment. On that account, norms prohib-

iting actions that are likely to elicit strong negative affect will be treated as distinctively wrong. Suffering in others triggers strong negative affect in the form of contagious distress, personal distress, and concern. As a result, norms prohibiting actions that cause suffering in others will count as Sentimental Rules. Furthermore, the responses to suffering are present in children by two years of age. Hence, children have these emotional responses well in place by the time they treat harm-based violations as distinctively wrong.