19 Cognitive Neuroscience and the Structure of the Moral Mind

Joshua Greene

(forthcoming in *Innateness and the Structure of the Mind, Vol. I.* Laurence, S., Carruthers, P. and Stich. S. eds., Oxford University Press.)

If you visit www.dictionary.com and type in the word "innate," this is what you'll get:

> *adj*
> 1.Possessed at birth; inborn.
> 2.Possessed as an essential characteristic; inherent.
> 3.Of or produced by the mind rather than learned through experience: *an innate knowledge of right and wrong*.

Of all the things in the world one might use to illustrate the concept of innateness, this dictionary offers *moral knowledge*. I find this amusing – the idea that someone who is not exactly sure what "innate" means would benefit from knowing that one of the most complex and least understood of human capacities could plausibly be described as "innate." And yet this choice, I suspect, is no accident. Our capacity for moral judgment, perhaps more than anything else, strikes people as both *within us* and *external to us*, as essentially human and at the same time possessing a mysterious external authority, like the voice of God or Nature calling us at once from within and beyond. But however obvious the reality of an innate capacity for moral judgment may be to theologians, lexicographers, and the like, it is not at all obvious from a scientific point of view, or even clear what such a capacity would amount to.

Any investigation into the possibility of an innate capacity for moral judgment must begin with what is known about moral psychology. Much of what we know comes from the developmental tradition, beginning with the work of Piaget (Piaget, 1965) and Kohlberg (Kohlberg, 1969). Some of the most compelling work on moral psychology has come from studies of the social behavior of our nearest living relatives, especially the great apes (de Waal, 1996; Flack and de Waal, 2000). Such

studies reveal what Flack and de Waal call the "building blocks" of human morality. Likewise, anthropologists (Shweder*, et al.*, 1997), evolutionary psychologists (Cosmides, 1989; Wright, 1994), and evolutionary game theorists (Axelrod, 1984; Sober and Wilson, 1998) have made other important contributions. Perhaps the most striking work of all has come from "Candid Camera"-style studies from within the social psychological tradition that dramatically illustrate the fragility and capriciousness of human morality (Milgram, 1974; Ross and Nisbett, 1991). All of these disciplines, however, treat the mind as a "black box," the operations of which are to be inferred from observable behavior. In contrast, the emerging discipline of cognitive neuroscience aims to go a level deeper, to open the mind's black box and thus understand its operations in physical terms. The aim of this chapter is to discuss neuro-cognitive work relevant to moral psychology and the proposition that innate factors make important contributions to moral judgment.

1 Lesion data

Imagine the following scenario. A woman is brought to the emergency room after sustaining a severe blow to the head. At first, and much to her doctors' surprise, her neurological function appears to be completely normal. And for the most part it is, but it soon becomes clear that she has acquired a bizarre disability. As a result of her accident, this woman can no longer play basketball. Her tennis game is still top notch, as is her golf swing, and so on. Only her basketball game has been compromised. Could such an accident really happen? Almost certainly not. The way the brain is organized, it is virtually impossible that something like a blow to the head could selectively destroy one's ability to play basketball and nothing else. This is because the neural machinery required to play basketball isn't sitting in one place, like a car's battery (Casebeer and Churchland, 2003). Instead, this machinery is distributed throughout the brain, and its various components are used in the performance of any number of other tasks.

While no one claims to have seen a case of acquired "abasketballia," there have been cases in which brain damage has appeared to rob individuals of their moral sensibilities in a strikingly selective way. By far, the most celebrated of such cases is that of Phineas Gage (Damasio, 1994), a Nineteenth Century railroad foreman who worked in Vermont. One fateful day, an accidental explosion sent a tamping iron through Gage's cheek and out the top of his head, destroying much of his medial prefrontal cortex. Gage not only survived the accident; at the time he

appeared to have emerged with all of his mental capacities intact. After a two-month recuperation period Gage was pronounced cured, but it was soon apparent that Gage was damaged. Before the accident he was admired by his colleagues for his industriousness and good character. After the accident, he became lawless. He wandered around, making trouble wherever he went, unable to hold down a steady job due to his anti-social behavior. For a long time no one understood why Gage's lesion had the profound but remarkably selective effect that it had.

More recent cases of patients with similar lesions have shed light on Gage's injury. Damasio and colleagues (Damasio, 1994) report on a patient named "Elliot" who suffered a brain tumor in roughly the same region that was destroyed in Gage. Like Gage, Elliot has maintained his ability to speak and reason about topics such as politics and economics. He scores above average on standard intelligence tests, including some designed to detect frontal lobe damage, and responds normally to standard tests of personality. However, his behavior, like Gage's, is not unaffected by his condition. While Elliot did not develop anti-social tendencies to the extent that Gage did, he, too, exhibits certain peculiar deficits, particularly in the social domain. A simple laboratory probe has helped reveal the subtle but dramatic nature of Elliot's deficits. When shown pictures of gory accidents or people about to drown in floods, Elliot reports having no emotional response but comments that he knows that he used to have strong emotional responses to such things. Intrigued by these reports, Damasio and colleagues employed a series of tests designed to assess the effects of Elliot's damage on his decision-making skills. They asked him, for example, whether or not he would steal if he needed money and to explain why or why not. His answers were like those of other people, citing the usual reasons for why one shouldn't commit such crimes. Saver and Damasio followed up this test with a series of five tests of moral/social judgment (Saver and Damasio, 1991). As before, Elliot performed normally or above average in each case. It became clear that Elliot's explicit knowledge of social and moral conventions was as good or better than most people's, and yet his personal life, like Gage's, has deteriorated rapidly as a result of his condition (although he does not seem to mind). Damasio attributes Elliot's real-life failures not to his inability to reason, but to his inability to integrate emotional responses into his practical judgments. "To know, but not to feel," says Damasio, is the essence of his predicament.

In a study of Elliot and four other patients with similar damage and deficits, Damasio and his colleagues observed a consistent failure to exhibit normal

electrodermal responses (a standard indication of emotional arousal) when these patients were presented with socially significant stimuli, though they responded normally to non-social, emotionally arousing stimuli (Damasio*, et al.*, 1990). A more recent study of patients like Elliot used the "Iowa gambling task" to study their decision-making skills (Bechara*, et al.*, 1996). In performing this task, patients like Elliot tend to make unwise, risky choices and fail to have normal electrodermal responses in anticipation to making those poor choices, suggesting, as predicted, that their failure to perform well in the gambling task is related to their emotional deficits. They can't *feel* their way through the problem.

While the subjects in the above studies exhibit "sociopathic behavior" as a result of their injuries, they are not "psychopaths." Most often they themselves, rather than others, are the victims of their poor decision-making. However, a more recent study (Anderson*, et al.*, 1999) of two subjects whose ventral, medial, and polar prefrontal cortices were damaged at an early age (three months and fifteen months) reveals a pattern of behavior that is characteristically psychopathic: lying, stealing, violence, and lack of remorse after committing such violations. These developmental patients, unlike Elliot and the like., exhibit more flagrantly anti-social behavior, presumably because they did not have the advantage of a lifetime of normal social experience involving normal emotional responses. Both patients perform fairly well on IQ tests and other standard cognitive measures and perform poorly on the Iowa gambling task, but unlike adult-onset patients their knowledge of social/moral norms is deficient. Their moral reasoning appears to be, in the terminology of Kohlberg, "preconventional," conducted from an egocentric perspective in which the purpose is to avoid punishment. Other tests show that they have a limited understanding of the social and emotional implications of decisions and fail to identify primary issues and generate appropriate responses to hypothetical social situations. Grattan and Eslinger (Grattan and Eslinger, 1992) report similar results concerning a different developmental-frontal patient. Thus, it appears that the brain regions compromised in these patients include structures crucial not only for online decision-making but also for the acquisition of social knowledge and dispositions toward normal social behavior.

What can we learn from these damaged individuals? In Gage – the legend if not the actual patient – we see a striking dissociation between "cognitive"[1] abilities

[1] The term "cognitive" has two uses. In some contexts, "cognitive" refers to information processing in a general. In other contexts, "cognitive" refers to a more narrow range of processes that contrast with

and moral sensibilities. Gage, once an esteemed man of character, is transformed by his accident into a scoundrel, with little to no observable damage to his "intellectual" faculties. A similar story emerges from Elliot's normal performance on questionnaire-type assays of his social/moral decision-making. Intellectually, or "cognitively," Elliot knows the right answers, but his real life social/moral decision-making is lacking. From this pattern of results, one might conclude that Gage, Elliot, and the like have suffered selective blows to their "morality centers." Other results, however, complicate this neat picture. Elliot and similar patients appear to have emotional deficits that are somewhat more general and that adversely affect their decision-making in non-social contexts as well as social ones (e.g. on the gambling task). And to further complicate matters, the developmental patients studied by Anderson and colleagues appear to have some "cognitive" deficits, although these deficits are closely related to social decision-making. Thus, what we observe in these patients is something less than selective damage to these individuals' moral judgment abilities, but something more than a general deficit in "reasoning" or "intelligence" or "judgment." In other words, these data suggest that there are dissociable cognitive systems that contribute asymmetrically to moral judgment but give us little reason to believe that there is a discrete faculty for moral judgment or a "morality module."[2] What these data do suggest is that there is an important dissociation between affective and "cognitive" contributions to social/moral decision-making and that the importance of the affective contributions has been underestimated by those who think of moral judgment primarily as a reasoning process (Haidt, 2001).

2 Anti-social behavior

The studies described above concern patients whose social behavior has been compromised by observable and relatively discrete brain lesions. There are, however, many cases of individuals who lack macroscopic brain damage and who exhibit pathological social behavior. These people fall into two categories: people with anti-social personality disorder (APD) and the subset of these individuals known

---

affective or emotional processes. Here, I reluctantly use the term "cognitive" with scare quotes to indicate the second meaning.

[2] There is a sizable literature reporting on patients with morally aberrant behavior resulting from frontal damage, and the cases discussed above are not necessarily representative (Grafman et al., 1996). I have chosen to focus on these cases because they involve what I take to be the most interesting dissociations between moral and other capacities.

as psychopaths. Anti-social personality disorder is a catch-all label for whatever it is that causes some people to habitually violate our more serious social norms, typically those that are codified in our legal system (DSM IV, 1994). Psychopaths not only engage in anti-social behavior, but exhibit a pathological degree of callousness, lack of empathy or emotional depth, and lack of genuine remorse for their anti-social actions (Hare, 1991). In more intuitive terms, the difference between APD and psychopathy is something like the difference between a hot-headed barroom brawler and a cold-blooded killer.

Psychopaths appear to be special in a number of ways (Blair, 2001). First, while the behavioral traits that are used to diagnose APD correlate with IQ and socio-economic status, the traits that are distinctive of psychopaths do not (Hare*, et al.*, 1991). Moreover, the behaviors associated with APD tend to decline with age, while the psychopath's distinctive social-emotional dysfunction holds steady (Harpur and Hare, 1994). The roots of psychopathic violence appear to be different from those of similarly violent non-psychopaths. In two ways, at least, their violence appears to be less contingent on environmental input. First, positive parenting strategies appear to influence the behavior of non-psychopaths, whereas psychopaths appear to be impervious in this regard (Wootton*, et al.*, 1997). Second, and probably not incidentally, the violence of psychopaths is more often instrumental rather than impulsive (Blair, 2001).

Experimental studies of psychopaths reveal further, more subtle differences between psychopaths and other individuals with APD. Psychopaths exhibit a lower level of tonic electrodermal activity and show weaker electrodermal responses to emotionally significant stimuli than normal individuals (Hare and Quinn, 1971). A more recent study (Blair*, et al.*, 1997) compares the electrodermal responses of psychopaths to a control group of criminals who, like the psychopathic individuals, were serving life sentences for murder or manslaughter. While the psychopaths resembled the other criminals in their responses to threatening stimuli (e.g. an image of a shark’s open mouth) and neutral stimuli (e.g. an image of a book), they showed significantly reduced electrodermal responses to distress cues (e.g. an image of a crying child’s face) relative to the control criminals, a fact consistent with the observation that psychopathic individuals appear to have a diminished capacity for emotional empathy. An earlier study (Blair, 1995) revealed that psychopaths, unlike ordinary criminals, have an impoverished appreciation of what is known as the "moral"/"conventional" distinction (Turiel, 1983). Most people believe that some

social rules may be modified by authority figures while others may not. For example, if the teacher says that it's okay to speak without raising one's hand ("conventional" violation), than it's okay to do so, but if the teacher says that it's okay to hit people ("moral" violation), then it's still not okay to hit people. Psychopaths seem to lack an intuitive understanding of this moral/conventional distinction, and it has been suggested that they perceive all social rules as mere rules (Blair, 1995). Finally, a recent study suggests that psychopathic murderers, unlike other murders and non-murdering psychopaths, fail to have normal negative associations with violence (Gray*, et al.*, 2003).

According to Blair (Blair*, et al.*, 1997), “The clinical and empirical picture of a psychopathic individual is of someone who has some form of emotional deficit." This conclusion is bolstered by the results of a recent neuroimaging study (Kiehl*, et al.*, 2001) in which psychopaths and control criminals processed emotionally salient words. The posterior cingulate gyrus, a region that exhibits increased activity during a variety of emotion-related tasks (Maddock, 1999), was less active in the psychopathic group than in the control subjects. At the same time, other regions were more active in psychopaths during this task, leading Khiel *et al.* to conclude that the psychopaths were using an alternative cognitive strategy to perform this task.

Thus, so far, a host of signs point to the importance of emotions in moral judgment (Haidt, 2001). In light of this, one might come to the conclusion that a psychopath, with his dearth of morally relevant emotion, is exactly what we're looking for – a human being "with everything – hold the morality." Indeed, Schmitt et al. (Schmitt*, et al.*, 1999) found that psychopaths performed normally on the Iowa gambling task, suggesting that their emotion-based decision-making deficits are not general, but rather related specifically to the social domain. As before, however, the empirical picture is not quite so simple, as psychopaths appear to have other things "held" as well. To begin, two studies, one of adult psychopaths (Mitchell*, et al.*, 2002) and one of children with psychopathic tendencies (Blair*, et al.*, 2001), found that psychopathic individuals do perform poorly on the Iowa gambling task. (These authors attribute the conflicting results to Schmitt et al.'s failure to use the original task directions, which emphasize the strategic nature of the task.) Moreover, there are several indications that psychopaths have deficits that extend well beyond their apparently stunted social-emotional responses. They respond abnormally to a number of "dry" cognitive tasks, both in terms of their behavior (Bernstein*, et al.*, 2000; Lapierre*, et al.*, 1995; Newman*, et al.*, 1997) and their

electrorencephalographic ("brainwave") responses (Kiehl*, et al.*, 1999a; Kiehl*, et al.*, 1999b; Kiehl*, et al.*, 2000). A common theme among these studies seems to be psychopaths' one-track-mindedness, their inability to inhibit prepotent responses and respond to peripheral cues.

The psychopathy literature sends mixed signals regarding the "impulsivity" of psychopaths. Psychopathic violence is has been described as "instrumental" rather than "reactive" (Blair, 2001). At the same time, however, some of the evidence described above suggests that psychopaths have a hard time inhibiting disadvantageous behavior, even during the performance of "dry" cognitive tasks. Compared to some anti-social individuals, psychopaths are "cool and collected," but a closer examination reveals that psychopaths have a kind of impulsivity or one-track-mindedness that subtly distinguishes them from normal individuals. The results of a neuroimaging study of "predatory" vs. "affective" murderers (Raine*, et al.*, 1998) gestures toward a synthesis. Raine *et al*. argue that excessive sub-cortical activity in the right hemisphere leads to violent impulses, but that "predatory" murderers, who unlike "affective" murderers exhibit normal levels of prefrontal activity, are better able to control these impulses. (In a more recent study (Raine*, et al.*, 2000), it was found that a sample of individuals diagnosed with APD (some of whom, however, may have been psychopaths) tended on average to have decreased prefrontal gray matter.) However, it's not clear how to reconcile the claim that "predatory" and "affective" murderers act on the same underlying impulses with the claim that psychopathic violence is "instrumental" rather than "impulsive."

In sum, psychopaths are not nature's controlled experiment with amorality. Psychopathy is a complicated syndrome that has subtle and not-so-subtle effects on a wide range of behaviors, including many behaviors that, superficially at least, have nothing to do with morality. At the same time, however, psychopathy appears to be a fairly specific syndrome. Psychopaths are not just people who are unusually anti-social. Using the proper methods, psychopaths are clearly distinguishable from others whose behavior is comparably anti-social, suggesting that the immoral behavior associated with psychopathy stems from the malformation of specific cognitive structures that make important contributions to moral judgment. Moreover, these structures seem to be rather "deep" in the sense that they are not well-defined by the concepts of ordinary experience and, more to the point, ordinary learning. Psychopaths do not appear to be people who have, through some unusual set of experiences, acquired unusual moral beliefs or values. Rather, they appear to have

an abnormal but stereotyped cognitive structure that affects a wide range of behaviors, from their willingness to kill to their inability to recall where on a screen a given word has appeared (Bernstein*, et al.*, 2000).

3 Neuroimaging studies of moral judgment and decision-making

Consider the following moral dilemma (the *trolley* dilemma (Foot, 1978; Thomson, 1986)): A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save these people is to hit a switch that will turn the trolley onto an alternate set of tracks where it will run over and kill one person instead of five. Is it okay to turn the trolley in order to save five people at the expense of one? Most people I've tested say that it is, and they tend to do so in a matter of seconds (Greene*, et al.*, 2001).

Now consider a slightly different dilemma (the *footbridge* dilemma (Thomson, 1986)): A runaway trolley threatens to kill five people as before, but this time you are standing next to a large stranger on a footbridge spanning the tracks, in between the oncoming trolley and the five people. The only way to save the five people is to push this stranger off the bridge and onto the tracks below. He will die as a result, but his body will stop the trolley from reaching the others. Is it okay to save the five people by pushing this stranger to his death? Most people I've tested say that it's not and, once again, they do so rather quickly.

These dilemmas were devised as part of a puzzle for moral philosophers (Foot, 1978; Thomson, 1986) by which the aim is to explain why it's okay to sacrifice one life to save five in the first case but not in the second case. Solving this puzzle has proven very difficult. While many attempts to provide a consistent, principled justification for these two intuitions have been made, the justifications offered are not at all obvious and are generally problematic. The fact that these intuitions are not easily justified gives rise to second puzzle, this time for moral psychologists: How do people know (or "know") to say "yes" to the *trolley* dilemma and "no" to the *footbridge* dilemma if there is no obvious, principled justification for doing so? If these conclusions aren't reached on the basis of some readily accessible moral principle, they must be made on the basis of some kind of intuition. But where do these intuitions come from?

To try to answer this question, my colleagues and I conducted an experiment in which subjects responded to these and other moral dilemmas while having their brains scanned (Greene*, et al.*, 2001). We hypothesized that the thought of pushing

someone to his death with one's bare hands is more emotionally salient than the thought of bringing about similar consequences by hitting a switch. More generally, we supposed that moral violations of an "up close and personal" nature, as in the footbridge case, are more emotionally salient than moral violations that are more impersonal, as in the trolley case, and that this difference in emotional response explains why people respond so differently to these two cases.

The rationale for this hypothesis is evolutionary. It is very likely that we humans have inherited many of our social instincts from our primate ancestors, among them instincts that rein in the tendencies of individuals to harm one another (de Waal, 1996; Flack and de Waal, 2000). These instincts are emotional, triggered by behaviors and other elicitors that were present in our ancestral environment. This environment did not include opportunities to harm other individuals using complicated, remote-acting machinery, but it did include opportunities to harm other individuals by pushing them into harms way (e.g. off a cliff or into a river). Thus, one might suppose that the sorts of basic, interpersonal violence that threatened our ancestors back then will "push our buttons" today in a way that peculiarly modern harms do not.

With all of this in mind, we operationalized the "personal"/"impersonal" distinction as follows: A moral violation is personal if it is (a) likely to cause serious bodily harm (b) to a particular person (c) in such a way that the harm does not result from the deflection of an existing threat onto a different party. (Cf. the “no new threat principle”(Thomson, 1986).) A moral violation is impersonal if it fails to meet these criteria. One can think of these criteria for personal harm in terms of ME HURT YOU and as delineating roughly those violations that a chimpanzee can appreciate. Condition (a) (HURT) picks out roughly those harms that a chimp can understand (e.g., assault vs. tax evasion). Condition (b) (YOU) requires that the victim be vivid as an individual. Finally, condition (c) (ME) captures the notion of “agency,” the idea that the action must spring in a vivid way from the agent’s will, must be “authored” rather than merely “edited” by the agent. Pushing someone in front of a trolley meets all three criteria and is therefore "personal," while diverting a trolley involves merely deflecting an existing threat, removing a crucial sense of “agency” and therefore making this violation "impersonal." Other moral dilemmas (about forty in all) were categorized using these criteria as well.

Before turning to the data, the evolutionary rationale for the "personal"/"impersonal" distinction requires a bit more elaboration. Emotional

responses may explain why people say "no" to the *footbridge* dilemma, but why do they say "yes" to the *trolley* dilemma? Here we must consider what's happened since we and our closest living relatives parted ways. We, unlike other species, have a well-developed capacity for general-purpose abstract reasoning, a capacity that can be used to think about anything one can name, including moral matters. Thus, one might suppose that when the heavy-duty, social-emotional instincts of our primate ancestors lay dormant, abstract reasoning has an opportunity to dominate. And, more specifically, one might suppose that in response to the trolley case, with its peculiarly modern method of violence, the powerful emotions that might otherwise say "No!" remain quiet, and a faint little “cognitive” voice can be heard: "Isn't it better to save five lives instead of one?"

That's a hypothesis. Is it true? And how can we tell? This hypothesis makes some strong predictions regarding what we should see in people’s brains while they are responding to personal and impersonal moral dilemmas. The contemplation of personal moral dilemmas like the footbridge case should produce increased neural activity in brain regions associated with emotional response and social cognition, while the contemplation of impersonal moral dilemmas should produce relatively greater activity in regions associated with "higher cognition." This is exactly what was observed (Greene*, et al.*, 2001). Contemplation of personal moral dilemmas produced relatively greater activity in two emotion-related areas, the posterior cingulate cortex (the region Kiehl *et al*. (2001) found to exhibit decreased emotion-related activity in psychopaths) and the medial prefrontal cortex (one of the areas damaged in both Gage (Damasio*, et al.*, 1994) and Elliot (Bechara*, et al.*, 1996)), as well as in the superior temporal sulcus, a region associated with various kinds of social cognition in humans and other primates (Allison*, et al.*, 2000). A more recent replication of these results using a larger pool of subjects has revealed the same effect in the amygdala, one of the primary emotion-related structures in the brain (Greene *et al*., 2004). At the same time, contemplation of impersonal moral dilemmas produced relatively greater neural activity in two classically "cognitive" brain areas associated with working memory function in the inferior parietal lobe and the dorsolateral prefrontal cortex.

This hypothesis also makes a prediction regarding people's reaction times. According to the view I've sketched, people tend to have emotional responses to personal moral violations that incline them to judge against performing those actions. That means that someone who judges a moral violation to be appropriate (e.g.

someone who says it's okay to push the man off the bridge in the *footbridge* case) will most likely have to override an emotional response in order to do it. That overriding process will take time, and thus we would expect that "yes" answers will take longer than "no" answers in response to personal moral dilemmas like the *footbridge* case. At the same time, we have no reason to predict a difference in reaction times between "yes" and "no" answers in response to impersonal moral dilemmas like the *trolley* case because there is, according to this model, no emotional response (or much less of one) to override in such cases. Here, too, the prediction holds. Trials in which the subject judged in favor of personal moral violations took significantly longer than trials in which the subject judged against them, but there was no comparable reaction time effect observed in response to impersonal moral violations (Greene*, et al.*, 2001).

Further results support this model as well. Above we contrasted the neural effects of contemplating "personal" vs. "impersonal" moral dilemmas. But what should we expect to see if we subdivide the personal moral dilemmas into two categories based on difficulty (i.e. based on reaction time)? Consider the following moral dilemma (the *crying baby* dilemma): It's wartime, and you and some of your fellow villagers are hiding from enemy soldiers in a basement. Your baby starts to cry, and you cover your baby's mouth to block the sound. If you remove your hand your baby will cry, the soldiers will hear, and they will find you and the others and kill everyone they find, including you and your baby. If you do not remove your hand, your baby will smother to death. Is it okay to smother your baby to death in order to save yourself and the other villagers? This is a very difficult question. Different people give different answers and nearly everyone takes a relatively long time to answer.

Here's a similar dilemma (the *infanticide* dilemma): You are a teenage girl who has become pregnant. By wearing baggy clothes and putting on weight you have managed to hide your pregnancy. One day during school, you start to go into labor. You rush to the locker room and give birth to the baby alone. You do not feel that you are ready to care for this child. Part of you wants to throw the baby in the garbage and pretend it never existed so that you can move on with your life. Is it okay to throw away your baby in order to move on with your life? Among the people we tested, at least, this is a very easy question. All of them say that it would be wrong to throw the baby away, and most do so very quickly.

What's going on in these two cases? My colleagues and I hypothesized as follows. In both cases there is a prepotent, negative emotional response to the personal violation in question, killing one's own baby. In the *crying baby* case, however, there are powerful, countervailing "cognitively" encoded considerations that push one toward smothering the baby. After all, the baby is going to die no matter what, and so you have nothing to lose (in terms of lives lost/saved) and much to gain by smothering it, awful as it is. In some people the emotional response ("Aaaahhhh!!! Don't do it!!!") dominates, and those people say "no." In other people, a "cognitive," cost-benefit analysis ("But you have nothing to gain, and so much to lose...") wins out, and those people say "yes."

What does this model predict that we'll see in the brain data when we compare cases like *crying baby* to cases like *infanticide*? First, this model supposes that cases like *crying baby* involve an increased level of "response conflict," i.e. conflict between competing representations for behavioral response. Thus, we should expect that difficult moral dilemmas like *crying baby* will produce increased activity in a brain region that is associated (albeit, controversially) with response conflict, the anterior cingulate cortex (Botvinick*, et al.*, 2001). Second, according to our model, the crucial difference between cases like *crying baby* and cases like *infanticide* is that dilemmas like *crying baby* involve "cognitive" considerations that compete with the prepotent, negative emotional response. Thus, we should expect to see increased activity in classically "cognitive" brain areas when we compare cases like *crying baby* to cases like *infanticide*, even though dilemmas like *crying baby* are personal moral dilemmas. As for emotion-related activity, the prediction is unclear. On the one hand, this model requires that emotional responses play an important role in both types of cases, leading to the prediction that there will be little observable difference in emotion-related areas of the brain. On the other hand, the type or level of emotional response that is involved in a protracted cognitive conflict as hypothesized to occur in *crying baby* may be differentiate it from the sort of quick emotional response that is hypothesized to be decisive in cases like *infanticide*. Thus, one might also expect to see some sort of additional emotion-related brain activity for the former cases.

The two clear predictions of this model have held (Greene*, et al.*, 2004). Comparing high reaction time personal moral dilemmas like *crying baby* to low reaction time personal moral dilemmas like *infanticide* revealed increased activity in the anterior cingulate (conflict) as well as the anterior dorsolateral prefrontal cortex

and the inferior parietal lobes, both classically "cognitive" brain regions (Greene, *et al*., 2004).

So far we have talked about neural activity correlated with the type of dilemma under consideration, but what about activity correlated with subjects' behavioral response? Does a brain look different when it's saying "yes" as compared to when it's saying "no" to questions like these? To answer this question we subdivided our dilemma set further by comparing the trials in which the subject says "yes" to difficult personal moral dilemmas like *crying baby* to trials in which the subject says "no" in response to such cases. Once again, we turn to the model for a prediction. If the cases in which people say "yes" are cases in which "cognition" wins, then we would expect to see more activity in the dorsolateral prefrontal cortex and/or parietal lobes in those cases. Likewise, if cases in which people say “no” are cases in which emotion wins, then we would expect to see more activity in emotion-related areas such as the posterior cingulate, medial prefrontal cortex, or the amygdala.

The first of these predictions held. “Cogntive” brain regions in both the anterior dorsolateral prefrontal cortex and in the inferior parietal lobes exhibited greater activity for trials in which personal moral violations were judged appropriate (“yes”) as compared to trials in which such violations were judged inappropriate (“no”). No brain regions, however, showed the opposite effect. Moreover, brain regions in the posterior cingulate and precuneus also exhibited greater levels of activity for “yes” answers, a surprise given our model. However, the overall pattern of activity in these sub-regions appears to be different from that of the larger brain regions described as being more active for personal as compared to impersonal moral judgments. A satisfying interpretation of these results will require a more fine-grained understanding of the functional neuro-anatomy in this poorly understood part of the brain (Greene *et al*, 2004).

The above results, taken together, provide support for the model sketched above according to which moral decisions are produced through an interaction between emotional and "cognitive" processes subserved by anatomically dissociable brain systems. Another recent brain imaging experiment further supports this model of moral judgment. Alan Sanfey, Jim Rilling, and colleages (Sanfey*, et al.*, 2003) conducted a brain imaging study of the Ultimatum Game in order to study the neural bases of people's sense of fairness. The Ultimatum Game works as follows: There is a sum of money, say $10, and the first player (the proposer) makes a proposal as to how to divide it up between herself and the other player. The second player, the

responder, can either accept the offer, in which case the money is divided as proposed, or reject the offer, in which case no one gets anything.

When both players are perfectly rational, purely motivated by financial self-interest, and these facts are known to the proposer, the outcome of the game is guaranteed. Because something is better than nothing, a rationally and financially self-interested responder will accept any non-zero offer. A rationally and financially self-interested proposer who knows this will therefore offer the responder as small a share of the total as possible, and the thus the proposer will get nearly all and the responder will get nearly none. This, however, is not what usually happens when people play the game, even when both players know that the game will only be played once. Proposers usually make offers that are fair (i.e. fifty-fifty split) or close to fair, and responders tend to reject offers that are more than a little unfair. Why does this happen?

The answer, once again, implicates emotion. This study reveals that unfair offers, as compared to fair offers, produce increased activity in the anterior insula, a brain region associated with anger, disgust, and autonomic arousal. Moreover, individuals' average levels of insula activity correlated positively with the percentage of offers they rejected and was weaker for trials in which the subject believed that the unfair offer was made by a computer program. But the insula is only part of the story. The anterior cingulate (the region mentioned above that is associated with response conflict) and the dorsolateral prefrontal cortex (one of the regions mentioned above that is associated with "higher cognition") were also more active in response to unfair offers. Moreover, for trials in which unfair offers were rejected, the level of activity in the insula tended to be higher than the level of activity in the dorsolateral prefrontal cortex, while the reverse was true of trials in which unfair offers were rejected. This result parallels very nicely the finding described above that increased (anterior) dorsolateral prefrontal cortex activity was observed when people judged personal moral violations to be appropriate (in spite of their emotions, according to our model).

Other neuroimaging results have shed light on the neural bases of moral judgment. Jorge Moll and colleagues have conduced two experiments using simple, morally significant sentences (e.g. "They hung and innocent.") (Moll*, et al.*, 2001; Moll*, et al.*, 2002a) and an experiment using morally significant pictures (e.g. pictures of poor abandoned children) (Moll*, et al.*, 2002b). These studies along with the ones described above implicate a wide range of brain areas in the processing of morally significant stimuli, with a fair amount of agreement (given the variety of tasks

employed in these studies) concerning which brain areas are the most important. In addition, many of the brain regions implicated by this handful of neuroimaging studies of moral cognition overlap with those implicated in neuroimaging studies of "theory of mind," the ability to represent others' mental states (Frith, 2001). (For a more detailed account of the neuroanatomy of moral judgment and its relation to related processes see Greene and Haidt (Greene and Haidt, 2002).) While many big questions remain unanswered, it is clear from these studies that there is no "moral center" in the brain, no "morality module." Moreover, moral judgment does not appear to be a function of "higher cognition," with a few emotional perturbations thrown in (Kohlberg, 1969). Nor do moral judgments appear to be driven entirely by emotional responses (Haidt, 2001). Rather, moral judgments appear to be produced by a complex network of brain areas subserving both emotional and "cognitive" processes (Greene and Haidt, 2002; Greene*, et al.*, 2001; Sanfey*, et al.*, 2003).

4 What in moral psychology is innate?

In extracting from the above discussion provisional answers to this question, it will be useful to distinguish between the form and content of moral thought. The *form* of moral thought concerns the nature of the cognitive processes that subserve moral thinking, which will surely be a function of the cognitive structures that are in place to carry out those processes. The *content* of moral thought concerns the nature of people's moral beliefs and attitudes, what they think of as right or wrong, good or bad, etc.. Thus, it could turn out that all humans have an innate tendency to think about right and wrong in a certain way without any tendency to agree on which things are right or wrong. With this distinction in mind, let us review the data presented above.

A number of themes emerge from studies of (1) patients with social behavioral problems stemming from brain injury, (2) psychopaths, and (3) the neural bases of moral judgment in normal individuals. Popular conceptions of moral psychology, bolstered by the legend of Phineas Gage and popular portrayals of psychopaths, encourage the belief that there must be a "moral center" in the brain. This does not appear to be the case. The lesion patients discussed above, both developmental and adult-onset, all have deficits that extend beyond the moral domain, as do the psychopaths that have been studied. Moreover, the results of brain imaging studies of moral judgment reveal that moral decision-making involves a diverse network of neural structures that are implicated in a wide range of other phenomena.

Nevertheless, the dissociations observed in pathological cases and in the moral thinking of normal individuals are telling. Most importantly, multiple sources of evidence point toward the existence of at least two relatively independent systems that contribute to moral judgment: (1) an affective system that (a) has its roots in primate social emotion and behavior; (b) is selectively damaged in psychopaths and certain patients with frontal brain lesions; and (c) is selectively triggered by personal moral violations, perceived unfairness, and, more generally, socially significant behaviors that existed in our ancestral environment. (2) a "cognitive" system that (a) is far more developed in humans than in other animals; (b) is selectively preserved in the aforementioned lesion patients and psychopaths; and (c) is not triggered in a stereotyped way by social stimuli. I have called these two different "systems," but they themselves are almost certainly composed of more specific subsystems. In the case of the affective system, it's subsystems are probably rather domain-specific, while the system that is responsible for “higher cognition,” though composed of subsystems with specific cognitive functions, is more flexible and more domain-general than the affective system and its subcomponents. Mixed in with what I've called the affective system are likely to be cognitive structures specifically dedicated to representing the mental states of others ("theory of mind") (Greene and Haidt, 2002).

What does this mean for the innateness of moral thought? It seems that the *form* of moral thought is highly dependent on the large-scale structure of the human mind. Cognitive neuroscience has made it increasingly clear that the mind/brain is composed of a set of interconnected modules. Modularity is generally associated with nativism, but some maintain that learning can give rise to modular structure, and in some cases this is certainly true (Elman*, et al.*, 1996; Shiffrin and Schneider, 1977). My opinion, however, is that large scale modular structure is unlikely to be produced without a great deal of specific biological adaptation to that end. Insofar as that is correct, the form of human moral thought is to a very great extent shaped by how the human mind happens to have evolved. In other words, our moral thinking is not the product of moral rules written onto a mental blank slate by experience. As the stark contrast between the *trolley* and *footbridge* dilemmas suggests, our moral judgment is greatly affected by the quirks in our cognitive design.

As for the *content* of human morality, there are good reasons to think that genes play an important role here as well. Many of our most basic pro-social tendencies are exhibited in other species such as the chimpanzee, suggesting that

such tendencies stem from shared genes (Flack and de Waal, 2000). Moreover, insofar as one can take modularity as evidence for innate structure, the fact that psychopaths exhibit relatively normal cognitive function along side dramatic deficits in emotional empathy suggests that normal empathic responses may depend on something like an innate "empathy module." (See also Tooby and Cosmides on innate motivation (Tooby and Cosmides, this volume).) Finally, the fact that psychopathic tendencies, unlike ordinary violent tendencies, appear to be unaffected by differences in parenting strategy (Wootton*, et al.*, 1997) and socioeconomic status (Hare*, et al.*, 1991) suggests that psychopathy may result from compromised genes.

So far I've argued that the form of human moral thought is importantly shaped by the innate structure of the human mind and that some basic, pro-social tendencies probably provide human morality with innate content. What about more ambitious versions of moral nativism? Might there be detailed moral principles written into the brain? People seem to "know" intuitvely that it's okay to hit the switch in the *trolley* case and that it's not okay to push the man in the *footbridge* case. Moreover, they seem to know these things without knowing how they know them, i.e. without any access to organizing principles. Such mysterious nuggets of apparent moral wisdom encourage the thought that somewhere, deep in our cognitive architecture, we're going to find the mother lode: an innate "moral grammar" (Harman, 2000; Rawls, 1971; Stich, 1993; Mikhail, 2000). (Or, more accurately, an innate "moral language" since such rules would have *content* as well as form.) Whether this more ambitious form of moral nativism will pan out remains to be seen. But already there is evidence suggesting that much of human moral judgment depends on dissociable "cognitive" and affective mechanisms that can compete with one another and that are not specifically dedicated to moral judgment (Greene and Haidt, 2002). It seems unlikely, then, that human moral judgment as a whole derives from a core moral competence that implements a set of normative-looking rules. Nevertheless, this motley picture of the moral mind is compatible with certain aspects of moral judgment's depending on cognitive structures that can be described as implementing something like a "grammar."

As noted above, I believe that the question of nativism in moral psychology commands attention because our moral thought is at once highly familiar and thoroughly alien. Our moral convictions are central to our humanity, and yet their origins are obscure, leading people to attribute them to supernatural forces, or their more naturalistic equivalents. For some, it seems, the idea of innate morality holds

the promise of *validation*. Our moral convictions, far from being the internalization of rules that we invented and taught one another, would be a gift from a universe wiser than ourselves. There is no doubt much wisdom in our moral instincts, but they, like all of nature's fabrications, will have their quirks and flaws. Those who seek redemption in the study of moral psychology are bound for disappointment. Nevertheless, we will surely benefit from a deeper understanding of human morality and its biological underpinnings.

Acknowledgements

Thanks to Andrea Heberlein for many helpful suggestions.

References

Allison, T., Puce, A. and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn Sci, 47*.

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D. and Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat Neurosci, 211*.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. American Psychiatric Association.

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.

Bechara, A., Tranel, D., Damasio, H. and Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cereb Cortex, 62*.

Bernstein, A., Newman, J. P., Wallace, J. F. and Luh, K. E. (2000). Left-hemisphere activation and deficient response modulation in psychopaths. *Psychol Sci, 115*.

Blair, R. J. (1995). A cognitive developmental approach to mortality: investigating the psychopath. *Cognition, 571*.

Blair, R. J., Jones, L., Clark, F. and Smith, M. (1997). The psychopathic individual: a lack of responsiveness to distress cues? *Psychophysiology, 342*.

Blair, R. J. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *J Neurol Neurosurg Psychiatry, 716*.

Blair, R. J., Colledge, E. and Mitchell, D. G. (2001). Somatic markers and response reversal: is there orbitofrontal cortex dysfunction in boys with psychopathic tendencies? *J Abnorm Child Psychol, 296*.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol Rev, 1083*.

Calder, A. J., Lawrence, A. D. and Young, A. W. (2001). Neuropsychology of fear and loathing. *Nat Rev Neurosci, 25*.

Casebeer, W. D. and Churchland, P. S. (2003). The Neural Mechanisms of Moral Cognition: A Multiple Aspect Approach to Moral Judgment and Decision-Making. *Biology and Philosophy, 18*

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 313*.

Critchley, H. D., Elliott, R., Mathias, C. J. and Dolan, R. J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *J Neurosci, 208*.

Damasio, A. R., Tranel, D. and Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behav Brain Res, 412*.

Damasio, A. R. (1994). *Descartes' error : emotion, reason, and the human brain*. G.P. Putnam.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J. and Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self- generated emotions. *Nat Neurosci, 310*.

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M. and Damasio, A. R. (1994). The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science, 2645162*.

de Waal, F. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Harvard University Press.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. and Plunkett, K. (1996). *Rethinking Innateness*. MIT Press.

Flack, J. C. and de Waal, F. B. M. (2000). 'Any animal whatever': Darwinian building blocks of morality in monkeys and apes. In Katz, L. D. (Eds.), *Evolutionary origins of morality*. Imprint Academic.

Fletcher, P. C., Frith, C. D., Baker, S. C., Shallice, T., Frackowiak, R. S. and Dolan, R. J. (1995). The mind's eye--precuneus activation in memory-related imagery. *Neuroimage, 23*.

Foot, P. (1978). The problem of abortion and the doctrine of double effect. In (Eds.), *Virtues and Vices*. Blackwell.

Frith, U. (2001). Mind blindness and the brain in autism. *Neuron, 326*.

Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H. R. and Salazar, A. M. (1996). Frontal lobe injuries, violence, and aggression: a report of the Vietnam Head Injury Study. *Neurology, 465*.

Grattan, L. M. and Eslinger, P. J. (1992). Long-term psychological consequences of childhood frontal lobe lesion in patient DT. *Brain Cogn, 201*.

Gray, N. S., MacCulloch, M. J., Smith, J., Morris, M. and Snowden, R. J. (2003). Forensic psychology: Violence viewed by psychopathic murderers. *Nature, 4236939*.

Greene, J. and Haidt, J. (2002). How (and where) does moral judgment work? *Trends Cogn Sci, 612*.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. and Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 2935537*.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M. and Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *unpublished manuscript,*

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*

Hare, R. D. and Quinn, M. J. (1971). Psychopathy and autonomic conditioning. *Journal of Abnormal Psychology, 77*

Hare, R. D. (1991). *The Hare psychopathy checklist-revised*. Multi-Health Systems.

Hare, R. D., Hart, S. D. and Harpur, T. J. (1991). Psychopathy and the DSM-IV criteria for antisocial personality disorder. *J Abnorm Psychol, 1003*.

Harman, G. (2000). Moral philosophy and linguistics. In (Eds.), *Explaining value*. Clarendon Press.

Harpur, T. J. and Hare, R. D. (1994). Assessment of psychopathy as a function of age. *J Abnorm Psychol, 1034*.

Kiehl, K. A., Hare, R. D., Liddle, P. F. and McDonald, J. J. (1999a). Reduced P300 responses in criminal psychopaths during a visual oddball task. *Biol Psychiatry, 4511*.

Kiehl, K. A., Hare, R. D., McDonald, J. J. and Brink, J. (1999b). Semantic and affective processing in psychopaths: an event-related potential (ERP) study. *Psychophysiology, 366*.

Kiehl, K. A., Smith, A. M., Hare, R. D. and Liddle, P. F. (2000). An event-related potential investigation of response inhibition in schizophrenia and psychopathy. *Biol Psychiatry, 483*.

Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J. and Liddle, P. F. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biol Psychiatry, 509*.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In Goslin, D. A. (Eds.), *Handbook of socialization theory and research*. Rand McNally.

Lapierre, D., Braun, C. M. and Hodgins, S. (1995). Ventral frontal deficits in psychopathy: neuropsychological test findings. *Neuropsychologia, 332*.

Maddock, R. J. (1999). The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. *Trends Neurosci, 227*.

Mikhail, J. (2000). Rawls' linguistic analogy: A study of the "generative grammar" model of moral theory described by John Rawls in A Theory of Justice. Cornell University, PhD Dissertation.

Milgram, S. (1974). *Obedience to authority; an experimental view*. Harper & Row.

Mitchell, D., Colledge, E., Leonard, A. and Blair, R. (2002). Risky decisions and response reversal: is there evidence of orbitofrontal cortex dysfunction in psychopathic individuals? *Neuropsychologia, 4012*.

Moll, J., Eslinger, P. J. and Oliveira-Souza, R. (2001). Frontopolar and anterior temporal cortex activation in a moral judgment task: preliminary functional MRI results in normal subjects. *Arq Neuropsiquiatr, 593-B*.

Moll, J., de Oliveira-Souza, R., Bramati, I. and Grafman, J. (2002a). Functional networks in emotional moral and nonmoral social judgments. *Neuroimage, 163 Pt 1*.

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A. and Pessoa, L. (2002b). The neural correlates of moral

sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *J Neurosci, 227*.

Newman, J. P., Schmitt, W. A. and Voss, W. D. (1997). The impact of motivationally neutral cues on psychopathic individuals: assessing the generality of the response modulation hypothesis. *J Abnorm Psychol, 1064*.

Nichols, S. (this volume). Innateness and moral psychology. In (Eds.),

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., Bullmore, E. T., Perrett, D. I., Rowland, D., Williams, S. C., Gray, J. A. and David, A. S. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature, 3896650*.

Raine, A., Meloy, J. R., Bihrle, S., Stoddard, J., LaCasse, L. and Buchsbaum, M. S. (1998). Reduced prefrontal and increased subcortical brain functioning assessed using positron emission tomography in predatory and affective murderers. *Behav Sci Law, 163*.

Raine, A., Lencz, T., Bihrle, S., LaCasse, L. and Colletti, P. (2000). Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Arch Gen Psychiatry, 572*.

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Ross, L. and Nisbett, R. E. (1991). *The person and the situation*. McGraw-Hill.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. and Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science, 3005626*.

Saver, J. L. and Damasio, A. R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia, 2912*.

Schmitt, W. A., Brinkley, C. A. and Newman, J. P. (1999). Testing Damasio's somatic marker hypothesis with psychopathic individuals: risk takers or risk averse? *J Abnorm Psychol, 1083*.

Shiffrin, R. M. and Schneider, W. (1977). Controlled and automaitc information processing: II. perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*

Shweder, R. A., Much, N. C., Mahapatra, M. and Park, L. (1997). The "big three" of morality (autonomy, community, and divinity), and the "big three" explanations of suffering. In Brandt, A. and Rozin, P. (Eds.), *Morality and Health*. Routledge.

Sober, E. and Wilson, D. S. (1998). *Unto others : the evolution and psychology of unselfish behavior*. Harvard University Press.

Stich, S. P. (1993). Moral philosophy and mental representation. In Hechtor, M., Nadel, L. and Michod, R. E. (Eds.), *Origin of Values*. Aldine de Gruyter.

Thomson, J. J. (1986). *Rights, restitution, and risk : essays, in moral theory*. Harvard University Press.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Wootton, J. M., Frick, P. J., Shelton, K. K. and Silverthorn, P. (1997). Ineffective parenting and childhood conduct problems: the moderating role of callous-unemotional traits. *J Consult Clin Psychol, 652*.

Wright, R. (1994). *The moral animal: evolutionary psychology and everyday life*. Pantheon.