Judgment and Decision Making, Vol. 3, No. 2, February 2008, pp. 121–139

The psychology of moral reasoning

Monica Bucciarelli*
Centro di Scienza Cognitiva and Dipartimento di Psicologia
University of Turin

Sangeet Khemlani and P. N. Johnson-Laird
Department of Psychology
Princeton University

Abstract

This article presents a theory of reasoning about moral propositions that is based on four fundamental principles. First, no simple criterion picks out propositions about morality from within the larger set of deontic propositions concerning what is permissible and impermissible in social relations, the law, games, and manners. Second, the mechanisms underlying emotions and deontic evaluations are independent and operate in parallel, and so some scenarios elicit emotions prior to moral evaluations, some elicit moral evaluations prior to emotions, and some elicit them at the same time. Third, deontic evaluations depend on inferences, either unconscious intuitions or conscious reasoning. Fourth, human beliefs about what is, and isn't, moral are neither complete nor consistent. The article marshals the evidence, which includes new studies, corroborating these principles, and discusses the relations between them and other current theories of moral reasoning.

Keywords: moral reasoning; deontic reasoning; intuitions; inferences; moral dilemmas.

1 Introduction

Is it morally wrong to take a paper-clip from your office? Is it morally wrong to steal money from your co-worker? Is it morally wrong to stab your employer? In the absence of mitigating circumstances, most individuals are likely to agree that all three actions are wrong, but that they increase in heinousness. Psychologists have studied moral evaluations and moral inferences for many years, but they have yet to converge on a single comprehensive theory of these processes (e.g., Blair, 1995; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Haidt, 2001; Hauser, 2006; Kohlberg, 1984; Piaget, 1965/1932). Our aim in the present article is to propose a new theory of moral reasoning, based on an account of inferences in general about permissible situations (Bucciarelli & Johnson-Laird, 2005), on a theory of emotions (Oatley & Johnson-Laird, 1996), and on an account of intuitions (Johnson-Laird, 2006). We begin with an outline of the principal psychological theories of how individuals make moral evaluations. We then describe the new theory. We present the evidence corroborating it, including some new experimental results. Finally, we consider the general nature of reasoning about moral propositions.

1.1 Psychological theories of moral reasoning

Psychologists have proposed various theories of moral reasoning, including those based on Piaget's "genetic epistemology" (see, e.g., Piaget, 1965/1932; Kohlberg, 1984). However, three current theories have been a source of ideas for us, and so in this section we outline their principal tenets. The first theory is due to Haidt (2001, 2007; see also Blair, 1995). Haidt proposes a "social-intuitionist" theory in which moral evaluations come from immediate intuitions and emotions in a process more akin to perception than reasoning. This view goes back to the Eighteenth century philosopher Hume (1978/1739), who wrote in his *Treatise of Human Nature*: "Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason. ... 'tis in vain to pretend, that morality

*For their helpful comments, we thank Emily Chakwin, Adele Goldberg, Sam Glucksberg, Geoffrey Goodwin, Louis Lee, Adam Moore, and Keith Oatley. We also thank Jon Baron, Nick Chater, Laurence Fiddick, Jonathan Haidt, and an anonymous reviewer, for the useful criticisms of an earlier version of the paper. Address: Monica Bucciarelli, Centro di Scienza Cognitiva and Dipartimento di Psicologia, Università di Torino, Turin 10123, Italy. Email: monica@psych.unito.it.

is discover'd only by a deduction of reason." By passions, Hume meant love, happiness, and other emotions. Haidt takes a similar view, because the social component of his theory postulates that conscious reasoning about moral issues comes only after intuitions about them, and that its role is solely to influence the intuitions of others. He takes moral intuitions to be "the sudden appearance in consciousness of a moral judgment, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of searching, weighing evidence, or inferring a conclusion" (Haidt, 2001, p. 818). Blair (1995, p.7) had proposed that it is the aversive feeling to transgressions – a feeling lacking on the part of psychopaths — that leads to the evaluation of transgressions as morally wrong. So, for Haidt (2001, p. 814), "moral intuitions (including moral emotions) come first and directly cause moral judgments." This account is of what happens "most of the time with most people": philosophers and others may be exceptions, and use prior conscious reasoning to evaluate issues in which they have no stake (Haidt, personal communication, 1–7–2008).

Haidt frames his theory as in opposition to Rationalism; and in the Eighteenth century, Hume's Empiricism was opposed by Rationalists, and in particular by Kant (1959/1785), who argued that a person's autonomy and self-governing rationality, not passion, was at the heart of morality. What makes individuals good is precisely that the moral law guides their decisions. Moral considerations are decisive, and, unlike other considerations, they are categorical, i.e., never to be qualified by circumstances. Hence, Kant's view of moral reasoning takes into account what, for him, is this unique characteristic of moral propositions. His categorical imperative asserts that individuals should act only in accordance with a maxim that they can at the same time will to be a universal law. This principle, as many modern philosophers agree, provides a four step procedure for moral decisions. First, you formulate a maxim capturing your reason for an action; second, you frame it as a universal principle for all rational agents; third, you assess whether a world based on this universal principle is conceivable; and, fourth, if it is, you ask yourself whether you would will the maxim to be a principle in this world. If you would, then your action is morally permissible (see, e.g., Hill, 1992). Suicide, for example, fails the third step, and so it is immoral. Lying for your own advantage fails at the fourth step, because a world in which everyone lived by the corresponding maxim is not one that you would intend. As these examples illustrate, the procedure for determining what is, and isn't permissible, depends on conscious reasoning about moral propositions.

The Rationalist tradition continues in modern thought, notably in Chomsky's accounts of natural language, and in his view that there is an innate universal grammar specifying all humanly possible languages (Chomsky, e.g., 1995). It contains a finite number of principles, and the settings of their parameters specify a finite but large number of different languages. The second main theory of moral reasoning likewise postulates an innate *moral* grammar (Hauser, 2006). The grammar is universal and equipped with a suite of principles and parameters for building moral systems. The principles are abstract, lacking specific content. Hauser writes (2006, p. 298): "Every newborn child could build a finite but large number of moral systems. When a child builds a particular moral system, it is because the local culture has set the parameters in a particular way. If you are born in Pakistan, your parameters are set in such a way that killing women who cheat on their husbands is not only permissible but obligatory, and the responsibility of family members." But, once the parameters are set, culture has little impact, and so it is no easier to acquire a second morality than a second language.

The resulting grammar automatically and unconsciously generates judgments of right and wrong for an infinite variety of acts and inactions. The judgments don't depend on conscious reasoning, and they don't depend on emotions, which couldn't make moral judgments. Instead, moral judgments trigger emotions, which are "downstream, pieces of psychology triggered by an unconscious moral judgment" (Hauser, p. 30–1; see also p. 156). In other words, emotions come after unconscious moral judgments. Mikhail (2007) defends the same view that a moral grammar yields rapid intuitive judgments with a high degree of certainty.

The theory is provocative, but not easy to test, because theorists have so far formulated only a few candidate rules for the grammar. But, Mikhail proposes two: the legal rule prohibiting intentional battery, and the legal rule of double effect, i.e., "an otherwise prohibited action, such as battery, that has both good and bad effects may be permissible if the prohibited act itself is not directly intended, the good but not the bad effects are directly intended, the good effects outweigh the bad effects, and no morally preferable alternative is available" (see also Foot, 1967; and Royzman & Baron, 2002).

Some evidence for moral grammars is that subtle differences in the framing of dilemmas can lead to different evaluations. Mikhail, for example, cites the contrast between these two versions of the well-known "trolley" dilemma:

1. A runaway trolley is about to run over and kill five people, but a bystander can throw a switch that will turn the trolley onto a side track, where it will kill only one person. Is it permissible to throw the switch?

2. A runaway trolley is about to run over and kill five people, but a bystander who is standing on a footbridge can shove a man in front of the train, saving the five peo-

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

ple but killing the man. Is it permissible to shove the man?

In one study, 90% of participants responded "yes" to dilemma 1, but only 10% responded "yes" to dilemma 2. The distinction between the two dilemmas, according to Mikhail, is between battery as a side effect (1) as in the law of double effect, and battery as a means (2), which is prohibited. An alternative explanation is that what matters is whether an action directly causes harm as in the second case, or only indirectly causes harm as in the first case (Royzman & Baron, 2002); and there are still other possibilities such as the nature of the intervention in the causal sequence (Waldmann & Dieterich, 2007).

Individuals who make these judgments can explain the basis of them in some cases, but they do not always allude to underlying principles (Cushman, Young, & Hauser, 2006), and so moral grammarians postulate that these intuitions reflect principles built into the moral grammar. Cushman et al. argue that one distinction between the two sorts of dilemma is between causing harm to a victim without using physical contact, and using physical contact to cause equivalent harm. The latter, they claim, is more blameworthy. Evolutionary psychologists, who postulate innate mental modules for reasoning, argue that pushing an individual in front of the trolley violates a rule in the social contract (Fiddick, Spampinato & Grafman, 2005).

The third theory of moral reasoning is due to Greene and his colleagues (see, e.g., Greene, et al., 2001). It amalgamates the Humean and Kantian traditions in a "dual process" account that posits two distinct ways in which individuals make moral evaluations. As Greene et al. remark (p. 2106): "Some moral dilemmas ... engage emotional processing to a greater extent than others, and these differences in emotional engagement affect people's judgments" (see also Nichols, 2002, for a comparable, though independent, theory). On Greene's account, the emotional reaction is to actions that are "up close and personal," and it is automatic. The idea of pushing a man in front of the trolley elicits an unpleasant emotion, and so individuals tend to evaluate the action as impermissible. In contrast, impersonal actions, such as the first version of the dilemma, elicit a reasoned response, and so it is permissible to throw the switch to divert the trolley, because it saves more lives. Reasoned responses, Greene proposes, are Utilitarian, that is, they are based on the doctrine that actions should yield the greatest good (or utility) to society (Bentham, 1996/1789; Mill, 1998/1863). Some psychologists have also argued that the Utilitarian doctrine provides a normative theory of morality (Baron, 2008, Ch. 16; Sunstein, 2005), but that moral heuristics — intuitions based on unconscious reasoning — often govern decisions, leading to deviations from the Utilitarian criterion.

Greene et al. (2001) reported an fMRI study of dilemmas that showed distinct brain mechanisms underlying the two sorts of reaction: personal dilemmas activated the limbic system that mediates basic emotions; impersonal dilemmas activated frontal regions underlying working memory and cognitive control. These investigators also reported that those who do decide that it is permissible to push the person in front of the trolley take longer to reach the decision, perhaps because they experience an emotion first, and reason afterwards. However, when Moore, Clark and Kane (2008) eliminated some confounds in the experimental materials, they failed to replicate this result. They observed that a measure of the processing capacity of working memory predicted judgments of permissibility in personal dilemmas for which harm was inevitable.

2 A theory of moral reasoning

All three theories in the previous section contain plausible components. But, a more comprehensive theory has to go beyond them. We now present such a theory, which incorporates some of their ideas in a synthesis leading to quite different empirical consequences. The theory presupposes an information-processing approach (Hunt, 1999), and it draws fundamental distinctions among emotions, intuitions, and conscious reasoning. We begin with an account of these distinctions, and of the different sorts of reasoning.

Reasoning or inference — we use the terms interchangeably — is any systematic mental process that constructs or evaluates implications from premises of some sort. Implications are either deductive or inductive (for this account, see, e.g., Johnson-Laird, 2006, Ch. 1). A deduction, or *valid* inference, yields a conclusion that must be true given that the premises are true. Any other sort of implication is an induction, e.g., an inference that isn't valid but that yields a conclusion likely to be true. Hence, a valid deduction never yields more information than is in its premises, whereas an induction, no matter how plausible its conclusion, goes beyond the information in its premises.

We can refine the categories of deduction and induction further, but for our purposes a more important and separate matter is that reasoning differs depending on whether individuals are conscious of its premises, and whether they are conscious of its conclusion. In common with other psychologists, we use the term *intuition* to refer to reasoning from unconscious premises, or from aspects of premises that are unconscious, to conscious conclusions. In contrast, we use *conscious* reasoning to refer to reasoning from conscious premises to conscious conclusions. Regardless of this contrast, the process of reasoning is itself largely unconscious.

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

The distinction between intuition and conscious reasoning is similar to "dual process" theories of reasoning advocated by many psychologists, including Reitman (1965), Johnson-Laird and Wason (1977), Evans and Over (1996), Sloman (1996), and Kahneman and Frederick (2005). These theories distinguish between rapid automatic inferences based on heuristics and slower conscious deliberations based on normative principles. For us, however, a key difference is that only conscious reasoning can make use of working memory to hold intermediate conclusions, and accordingly reason in a recursive way (Johnson-Laird, 2006, p. 69): primitive recursion, by definition, calls for a memory of the results of intermediate computations (Hopcroft & Ulmann, 1979). The following example illustrates this point:

Everyone is prejudiced against prejudiced people.
Anne is prejudiced against Beth.
Does it follow that Chuck is prejudiced against Di?

Intuition says: *no*, because nothing has been asserted about Chuck or Di. But, conscious reasoning allows us to make the correct chain of inferences. Because Anne is prejudiced against Beth, it follows from the first premise that everyone is prejudiced against Anne. Hence, Di is prejudiced against Anne. So, Di is prejudiced, and it follows from the first premise again that everyone is prejudiced against her. And that includes Chuck. So, Chuck *is* prejudiced against Di. The non-recursive processes of intuition cannot make this inference, but when we deliberate about it consciously, we grasp its validity (Cherubini & Johnson-Laird, 2004). Conscious reasoning therefore has a greater computational power than unconscious reasoning, and so it can on occasion overrule our intuitions.

2.1 Emotions and morals

Emotions are created by cognitive evaluations, which can be rudimentary and unconscious or complex and conscious. Emotional signals help you to co-ordinate your multiple goals and plans, given the constraints of time pressure and of your finite intellectual resources. They are more rapid than conscious reasoning, because they make no demands on your working memory. When you experience certain emotions, you may, or may not, know their cause. You can be happy with someone because she charmed you; but you can be happy for reasons that you do not know. On one account (Oatley & Johnson-Laird, e.g., 1996), only basic emotions, such as happiness, sadness, anger, and anxiety, can originate in unconscious evaluations. Emotions such as desire and disgust can be experienced only in relation to a known object. And complex emotions, such as jealousy and empathy, can be experienced only with a consciousness of their causes. Indeed, this consciousness elicits the emotion. Yet, in all cases, whether or not the cause is conscious, the mental *transition* to an emotion is unconscious and largely, if not totally, beyond control. One corollary is that some individuals may have unwanted basic emotions that are so prevalent and extreme that they suffer from a psychological illness (Johnson-Laird, Mancini, & Gangemi, 2006).

We now turn to the first question that concerns moral reasoning:

2.2 What are moral propositions?

The answer is that they are a sort of *deontic* proposition, and deontic propositions concern what you may, should, and should not do or else leave undone. Deontic propositions, however, often concern matters that have nothing to do with morality. In Western culture, you shouldn't eat peas with your knife. The offence is not to morals, but to manners. In a game of table tennis, you should start your service with the ball resting on the open palm of your stationary free hand. The obligation is not in itself a moral one, but occasioned by the laws of the game. Theories sometimes posit that there is a special sort of mechanism for *moral* reasoning. And so a prerequisite for these theories is to delineate those deontic propositions that concern moral issues, because the mechanism does not apply to other deontic matters, such as the conventions of table manners or the rules of table tennis.

Rationalists suggest that the truth of moral propositions, unlike those of etiquette or games, is not a matter of preference but of reason. Kant (1959/1785) himself drew a distinction between moral imperatives, which are good in themselves regardless of one's self interest, and other "hypothetical" imperatives, which are means to something else. A moral action is accordingly one that should be carried out for its own sake. There are several problems in treating this claim as a putative criterion for moral propositions. One difficulty is that it is not obvious how to assess whether or not an action should be carried out for its own sake, and is not in the agent's self-interest. Another difficulty, as Foot (1972) has pointed out, is that Kant's constraint exists for many conventions that are not matters of morality: regardless of your desires, you should play a let if your serve in tennis touches the net. Much the same argument can be made against the view that only morality provides reasons, or a rational basis, for action. There are also reasons for adopting conventions of etiquette and rules of games.

The philosopher, the late Richard M. Hare argued in a series of publications that three criteria govern moral propositions: such propositions are universal, applying to everyone for whom their preconditions hold; they are prescriptive in that they don't describe facts but rather tell you what to do or not to do; and they are evaluative in that they tell you what is right and wrong (see, e.g.,

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

Hare, 1981). These conditions seem to apply to all moral propositions, but, in our view, they also apply to other deontic propositions. Consider, for instance, the proposition about how to serve in table tennis. This proposition satisfies all three of Hare's criteria: it is universal, prescriptive, and evaluative. One counter-argument is that conventions, such as the rule for serving in table tennis, can become moral issues, depending on the attitudes of those applying them. Another counter-argument is that matters of convention can be altered by a voluntary decision. The authorities governing table tennis can, and do, change the laws of the game. In contrast, moral laws are supposed to be immutable. Indeed, Kantians argue that they are *categorical* imperatives. But, again as Foot (1972) has argued, the imperatives of etiquette can be just as categorical as those of morality, so that for her being categorical or immutable fails to demarcate moral propositions. Before we formulate our own view on a criterion for moral propositions, we will consider what psychologists have had to say about the matter.

Psychologists have proposed various bases underlying children's capacity to distinguish between moral and other sorts of deontic proposition. Turiel and his colleagues argue that moral concepts concern welfare, justice, and rights, whereas social conventions concern acceptable behaviors that help to coordinate human interactions (e.g., Wainryb & Turiel, 1993, pp. 209–10). But, such a distinction seems to us to be partly circular, because the notions of justice and rights are themselves moral notions. Even so, the distinction fails in many cases. Consider, for example, a person with a cold who sneezes over someone else. The action is a violation of a person's welfare, but for most of us it is grossly ill-mannered rather than a moral violation. These authors also observed that children judge actions in the moral domain independently from whether there are rules governing these actions, e.g., stealing is wrong whether or not there is a rule about it; whereas the acceptability of conventional acts depends on the existence of a rule, e.g., you should wear a school uniform given that there is a rule to that effect. Could this distinction serve as a criterion to demarcate moral propositions from other deontic ones? Alas, some moral issues arise only in the context of rules, e.g., whether students who take their notes into an examination are immoral cheats depends entirely on the rules of the examination. Conversely, some social conventions apply even in the absence of rules, e.g., don't sneeze over other people.

Blair (e.g., 1995) derives the criterion for moral propositions from inhibition against violence amongst conspecifics. But, as Nichols (2002) has pointed out, such an inhibition, or the experience of aversion in the face of transgressions, cannot in itself yield a *moral* evaluation. Another putative criterion is that only moral transgressions merit punishment (cf. Davidson, Turiel, & Black, 1983). But, many immoral acts, such as a failure to keep a promise, hardly warrant punishment; and the criterion is plainly useless in picking out morally good actions.

At first sight, the following criterion seems promising: morality concerns "[the] rightness or wrongness of acts that knowingly cause harm to people other than the agent" (Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006). But, this criterion fails too, because many acts that agents carry out on themselves have been, or are, considered moral issues, e.g., suicide, substance abuse, self abuse. Conversely, not all acts that knowingly cause harm to other people are matters of morality. When you review a paper and reject it for publication, you are liable to hurt the author of the paper, and you may knowingly do so. Yet, that in itself does not raise a moral issue. Could emotion be the criterion? On the whole, we tend to have stronger emotions to moral lapses than to lapses of convention. Yet, emotions cannot demarcate moral propositions (cf. Nichols, 2002). Certain lapses in etiquette are more disgusting than the theft of a paperclip. Moreover, not for the first time have psychologists focused on the bad news: aversion to violence, disgust, and the need for punishment, tell us little about how we decide that an action is morally *good*.

In the light of the preceding analysis, the first assumption of our theory is:

1. The principle of moral indefinability: No simple principled way exists to tell from a proposition alone whether or not it concerns a moral issue as opposed to some other sort of deontic matter.

A simple criterion that a proposition concerns *deontic* matters is that it refers to what is permissible or not, or to what is obligatory or not, e.g., "you shouldn't eat so much". The principle of moral indefinability states that it is difficulty to pick out from within deontic propositions all and only those that concern morality. The decision depends in many cases on the attitudes of those individuals who are evaluating the proposition.

Of course, it doesn't follow that there is no domain of moral propositions, or that you cannot recognize instances of moral propositions and instances of non-moral propositions. You can. Euthanasia may or may not be immoral, but there is no doubt that it is a moral issue in our culture, whereas whether or not one eats peas with a knife is not. The problem is that even within a single culture, such as ours, no clear boundary exists between moral and non-moral propositions. Is smoking a moral issue? Is eating too much a moral issue? Is egotistical discourse a moral issue? The answers to these questions are not obvious, but it is clear that one shouldn't smoke, eat too much, or talk excessively about oneself. Each of these propositions is deontic, but whether or not they

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

are moral propositions is unclear. So, how do you recognize certain propositions as concerning moral issues? You have to learn which issues are moral ones in your society, and from this knowledge you can also make inferential extrapolations, but, as we have illustrated, the boundaries are not clear cut. Obviously, the principle of indefinability would be false if there were a simple way to demarcate all and only the moral propositions within the broader category of deontic propositions.

Readers might wonder why indefinability matters, and whether it tells them anything of interest. It is pertinent to the hypothesis that a special and dedicated mechanism exists for moral reasoning. If so, there must be a way for the mind to identify those propositions — the moral ones — to which the mechanism applies. But, if no simple criterion exists to pick out these propositions from within the wider set of deontic propositions, it is plausible that moral reasoning is just normal reasoning about deontic propositions that happen to concern morality. And we can invoke a single mechanism that copes with all deontic reasoning (Bucciarelli & Johnson-Laird, 2005).

Emotions are evolutionarily very much older than moral and deontic principles: all social mammals appear to have basic emotions (cf. De Waal, 1996). Likewise, you experience emotions in many circumstances that have no moral or deontic components whatsoever, e.g., when you listen to music. Conversely, when you determine that a trivial infringement is deontically wrong, you may not experience any emotional reaction, e.g., when you decide that it is wrong to steal a paperclip. Hence, the next assumption of the present theory is as follows:

2. The principle of independent systems: Emotions and deontic evaluations are based on independent systems operating in parallel.

Consider this brief scenario:

> A couple's two sons stabbed them and left them to bleed to death in order to inherit their money.

It describes an event that is both horrifying and immoral, and you are likely to experience the emotion and to make the moral evaluation. In general, you may feel antipathetic emotions of anger, or revulsion, and disapprove of acts that are morally bad, such as instances of violence, dishonesty, or cowardice. You may feel a sympathetic emotion of happiness and approve acts that are morally good, such as instances of generosity, self-sacrifice, or courage. As we pointed out, some theories imply that emotions can contribute to moral evaluations (Haidt, 2001; Greene et al., 2001), and some theories imply that moral evaluations can contribute to emotions (Hauser, 2006). According to the principle of independent systems, neither view is quite right. Instead, some situations should elicit an emotional response prior to a moral evaluation: they are "emotion prevalent"; some should elicit a moral evaluation prior to an emotional response: they are "evaluation prevalent"; and some should elicit the two reactions at the same time: they are "neutral in prevalence". This hypothesis would therefore be false if everyone had a uniform tendency to experience emotions prior to moral evaluations, or vice versa.

2.3 Deontic reasoning

The principle of moral indefinability suggests that no unique inferential mechanisms exist for dealing with moral propositions. If so, conscious reasoning about moral propositions must depend on the same process that underlies any sort of deontic reasoning. Logicians have developed deontic logics based on the two central concepts of obligation and permissibility, which can be defined in terms of one another: If you're *obligated* to leave, then it's not permissible for you not to leave. Likewise, if you're *permitted* to leave, then you're not obligated not to leave. Evidence presented elsewhere, however, supports the theory that deontic reasoning depends, not on logical rules of inference, but on mental models instead (Bucciarelli & Johnson-Laird, 2005). This "model" theory postulates that possibilities are central to reasoning, and that deontic propositions concern deontic possibilities, i.e., permissible states. Each model of a deontic proposition represents either a permissible state or in rarer cases a state that is not permissible. If one action is common to all models, which represent what is permissible, then it is obligatory. Some deontic propositions are categorical, such as: *thou shalt not kill*, but many propositions state a relation between possible and permissible states: if your serve in tennis hits the net cord then you must serve again.

A crucial prediction of the model theory is illustrated in the following problem:

> You are permitted to carry out only one of the following two actions:
>
> Action 1: Take the apple or the orange, or both.
>
> Action 2: Take the pear or the orange, or both.
>
> Are you permitted to take the orange?

The mental models of action 1 represent what it is permissible to take: you can take the apple, you can take the orange, or you can take both of them. They support the conclusion that you are permitted to take the orange. (If you consider the alternative action 2, its mental models support the same conclusion.) Hence, if you rely on mental models, then you will respond, "yes", to the question in the problem. However, the response is an illusion. If you took the orange then you would carry out both action 1 and action 2, contrary to the rubric that you are

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

permitted to carry out only one of them. Unlike mental models, the *complete* models of the problem take into account that when one action is permissible the other is not permissible. These models show that two states are permissible: either you take the apple alone, or else you take the pear alone. Hence, the correct response is that it is *not* permissible to take the orange. Experiments have shown that intelligent adults tend to succumb to such illusions, but to reason correctly about comparable problems for which the failure to think about what is impermissible does not lead to error (Bucciarelli & Johnson-Laird, 2005). This result is crucial because only the model theory among current proposals predicts it.

Intuitions about moral issues should depend on a general deontic mechanism. They have unconscious premises, and so if you are asked for the grounds for an intuition, you are dumbfounded. You hear a piece of piano music, say, and immediately have the intuition that it is by Debussy. You may well be right even if you have never heard the particular piece before. Yet, it may be quite impossible for you to say what it is about the music that elicits the inference. In a similar way, as Haidt (2001) has shown, you can have a moral intuition, say, that incest is wrong, but be dumbfounded if someone asks you why. You might be similarly dumbfounded if someone asks you why you shouldn't eat peas with your knife.

Pat sees that a newspaper has been lying outside her neighbor's front door all day, and so she takes it. Is that right or wrong? You are likely to say that it is wrong: you make a simple conscious inference from the premise that stealing is wrong. But, why is stealing wrong? You may cite the Ten Commandments. You may frame a philosophical answer based on an analysis of property (e.g., Miller & Johnson-Laird, 1976, p. 558–562). Or, once again you may be dumbfounded. But, whatever response you make, your judgment that Pat was wrong to take the newspaper is likely to depend on conscious reasoning from the premise that stealing is wrong.

The Humean thesis that a moral evaluation is based solely on an emotional reaction depends, in our view, either on a skeptical and impoverished view of reasoning (see Hume, 1978/1739) or on positing an inferential mechanism within the emotional system. A step in the latter direction is the hypothesis that a system of emotional appraisals forbids actions with the semantic structure of: *me hurt you* (Greene, Nystrom, Engell, Darley, & Cohen, 2004). The present theory, however, rests on an alternative assumption:

3. The principle of deontic reasoning: all deontic evaluations including those concerning matters of morality depend on inferences, either unconscious intuitions or conscious reasoning.

No contemporary theorist doubts that humans can make inferences about deontic matters, and several authors allow that individuals both have intuitions and reason consciously about moral issues (Cushman et al., 2006; Koenigs et al., 2007; Pizarro, Uhlmann, & Bloom, 2003). According to the social-intuitionist theory, however, conscious reasoning does not yield moral evaluations, which are solely a result of intuitions and moral emotions (Haidt, 2001). Hence, a crucial issue is whether clear cases occur in which individuals, apart from philosophers or other experts, reason consciously in order to make a moral evaluation (Wheatley & Haidt, 2005). No study in the literature appears to have established unequivocally a prior role for reasoning. As Cushman et al. (2006, p. 1087) remark: "A task for future studies is to design methodologies that provide strong evidence in favor of consciously reasoned moral judgments." The principle of deontic reasoning would be false if no moral evaluations ever depended on conscious reasoning (pace Haidt, 2001), or else if no moral evaluations ever depended on intuitions (pace Kant, 1959/1785).

2.4 Moral inconsistency

Everyday beliefs are often inconsistent, and you get along with these inconsistencies in part because their detection is computationally intractable and in part because you tend to rely on separate sets of beliefs in separate contexts (see Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000). An example of an inconsistency occurs in your thinking about causation. On the one hand, you assume that you can intervene to initiate a causal chain. You throw a switch, for example, and the light comes on. On the other hand, you assume that every event has a cause. The screen on your television set suddenly goes black, and, like many viewers of the final episode of *The Sopranos*, you infer that something has gone wrong with the set. Yet, if every event has a cause, you didn't initiate a causal chain when you threw the light switch, because your action, in turn, had a cause. This sort of inconsistency has led some commentators to conclude that there is no such thing as cause and effect (Salsburg, 2001, p. 185–6). Yet, causation is so deeply embedded in the meanings of words that this view is too drastic (see, e.g., Miller & Johnson-Laird, 1976).

Inconsistencies also occur in deontic systems. For example, despite the best conscious inferences of lawyers, legal systems often contain them. Suber (1990) has pointed out many examples, and quotes an English judge, Lord Halsbury to the following effect: "...the law is not always logical at all". Moral beliefs have not had the advantages (or disadvantages) of legal scrutiny, and so the final assumption of our theory is as follows:

4. The principle of moral inconsistency: the beliefs that are the basis of moral intuitions and conscious moral reasoning are neither complete nor consistent.

We define a *logical* system of morals as consisting of a set of consistent moral principles (axioms) and a method of valid reasoning. It will yield moral evaluations, such as that Pat was wrong to steal the newspaper, but it will fail to cover certain eventualities if the principles are incomplete. What a logical system cannot yield, however, are inconsistencies or conflicts: it cannot yield a case for both the permissibility and the impermissibility of an action, such as stealing a newspaper. A "grammar" in Chomsky's (1995) sense also precludes inconsistencies: a string of words cannot be both grammatical and ungrammatical according to the rules of a grammar. A moral grammar may fail to cover all eventualities, and it won't deliver an evaluation when key information about a situation is unknown, but it should be a logical system and not yield conflicts in which an action is both permissible and impermissible. In contrast, the principle of moral inconsistency predicts that individuals should encounter irresolvable moral conflicts from time to time. If not, the principle is false.

In summary, the principle of the indefinability of moral propositions renders rather implausible any theory that proposes a special mechanism for moral reasoning. If no simple way exists to pick out those situations to which the mechanism should apply, it may well be that there is no special mechanism. The implication is that moral emotions and moral reasoning may well be normal emotions and normal reasoning, which happen to concern moral matters. According to the principle of independent systems, the mechanisms underlying emotions are independent from those underlying deontic evaluations. They can influence each other, but the influence can flow in either direction. The principle of deontic reasoning implies that all deontic evaluations, including moral intuitions, depend on inferences. And the principle of moral inconsistency predicts the occurrence of inconsistencies in moral evaluations. We now turn to the evidence corroborating these principles.

3 Evidence for independent systems

3.1 Experiment 1

When you read a scenario, such as our earlier example of the sons who murdered their parents, according to Hauser (2006) your first reaction is a moral intuition and your emotional response comes later. Haidt (2001) allows that you first experience a moral intuition perhaps accompanied with an emotion. It is not clear how you should react according to Greene et al. (2004) because your emotion and your evaluation are unlikely to conflict. In contrast to these accounts, the principle of independent systems predicts that some scenarios are likely to elicit an emotion first: they are, as we remarked earlier, emotion prevalent. Other scenarios are likely to elicit a moral intuition first: they are evaluation prevalent. And still other scenarios may show no particular bias either way: they are neutral in prevalence. As an initial test of this prediction, and in order to develop the materials for a study of latencies, we carried out an experiment using a simple procedure in which the participants' task was to read a one-sentence scenario and to report which experience they had first, an emotional or a moral reaction, and then to rate the strength of both these reactions.

3.1.1 Method

Forty-seven students (46 females and 1 male; mean age 22 years) at Turin University took part as a group in the experiment for course credit. They evaluated 40 scenarios describing various moral and immoral actions. For each scenario, they wrote down whether their first reaction was emotional or moral, and they then rated the strength of each of these reactions on separate 5-point scales. We devised 20 sentences describing morally good events, and 20 sentences describing immoral events. The morally good events concerned such matters as telling the truth, taking care of children, helping others, marital fidelity, generosity, and kindness to animals. A typical example is:

> A woman donated one of her kidneys to a friend of hers who was suffering from a diseased kidney and, as a result saved him from a certain death.

The immoral events concerned such matters as violence towards others, cannibalism, robbery, incest, cruelty to children, maltreating animals, cheating others, bribery, and sexual abuse. A typical example is:

> A violent bully terrorized the playground and beat up a younger girl with a hammer for no apparent reason.

The experiment was carried out in Italian, and the Italian versions of the sentences were matched for number of syllables (a mean of 44.5 for the moral descriptions and of 44.4 for the immoral ones).

Each scenario was presented on a separate page of a booklet followed by a question: Which did you experience first: an emotional reaction or a moral reaction? Beneath this question, was the instruction: Assign a score to your emotional reaction on a five-point scale (put an "X" on the scale). A Likert scale was printed below this

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

instruction, and it ranged from 1 labeled, "Very strong bad emotion," through a mid-point labeled "50:50" to 5 labeled, "Very strong good emotion". A similar instruction asked the participants to assign a score on an analogous five-point scale for the moral reaction, running from 1 labeled, "Very strong negative evaluation" through the mid-point to 5 labeled, "Very strong positive evaluation". The booklets were assembled with the pages in different random orders.

3.1.2 Results and discussion

The morally good scenarios had mean ratings of 4.19 for emotion and 4.19 for morality, and the immoral scenarios had mean ratings of 1.49 for emotion and 1.32 for morality, where 1 was the "bad" end of both scales and 5 was the "good" end. Not surprisingly, the morally good scenarios had higher ratings on both the emotion and moral scales than the immoral scenarios (Wilcoxon tests on the differences over the 47 participants, z = 3.92 and 3.93, $p < .0001$ in both cases). The ratings of the strength of the moral and emotional reactions were highly correlated (Spearman's rho = .9, $p< .01$). In order to test whether the participants tended to show a consensus about which reaction came first, and to help us to classify the scenarios for the next experiment, we adopted a simple criterion for a consensus: any scenario in which 30 or more of the 47 participants agreed about which came first, emotion or evaluation, counted as an instance of a consensus, because such a bias is significant on Binomial test ($p < .05$, one tailed). On this basis, 19 out of the 40 scenarios (10 moral and 9 immoral ones) were emotion prevalent, and 8 scenarios (4 moral and 4 immoral) were evaluation prevalent. Moreover, both the emotion-prevalent and the evaluation-prevalent scenarios were significantly more numerous than the 2 out of 40 expected to be significant (at .05) by chance ($p< .001$ for both).

The scenario with the greatest emotion prevalence was:

> Two friends, although they lived in different countries, always met up to celebrate each other's birthday.

With hindsight, its moral content is good but slight, and 46 out of the 47 participants reported that they had an emotional reaction first. The scenario with the greatest evaluation prevalence was:

> A woman told deliberate lies to cause the imprisonment of a person who had committed no crime.

For this scenario, 41 out of the 47 participants reported that they had an evaluation first. We postpone until the discussion of the next experiment what makes a scenario emotion or evaluation prevalent. Of course, the fact that participants tend to agree about which came first, the emotion or the evaluation, is no guarantee that they were right. Introspective reports are notoriously unreliable about certain aspects of mental life, but suppose that they were accurate in this case, what then? One implication is that individuals should show the same difference in the latencies of their answers to questions about emotions and evaluations. Experiment 2 tested this prediction.

3.2 Experiment 2

Individuals should be faster to answer questions about their emotions on reading an emotion prevalent scenario, but faster to answer questions about their moral evaluations on reading an evaluation-prevalent scenario. The experiment tested this prediction for scenarios from the first experiment.

3.2.1 Method

54 undergraduates at Turin University (46 females and 8 males; mean age 25 years) took part in the experiment for course credit. They were tested individually in a computer-controlled experiment. The task consisted of 24 trials in which they read a scenario and then responded to a single question, which was presented at the start of each trial. There were three sorts of questions presented in three blocks of eight trials each: an emotion question (does it make you feel good or bad?), a moral question (is it right or wrong?), and a consequential question (should it be punished or rewarded?). The order of the three blocks was counterbalanced in the six possible orders over the participants.

The 24 trials consisted of scenarios from Experiment 1: twelve were emotion prevalent, eight were evaluation prevalent, and four were neutral in prevalence. The scenarios were assigned to the blocks at random in three ways, with the constraint that each scenario occurred equally often in each sort of block.

The participants were told to imagine that they were responding to items in the news, and that they would judge an item in terms of their emotional reaction, their moral reaction, or whether the protagonist should be punished or rewarded. They were not told that their responses would be timed, but instead that there was no time limit. Once the participants had understood the task, they proceeded to the experiment. The computer timed the interval from the onset of the scenario until the participant responded, and so the latency of a response included the time to read and to understand the scenario, and the time to answer the question. The computer presented the response options over the relevant keys.

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

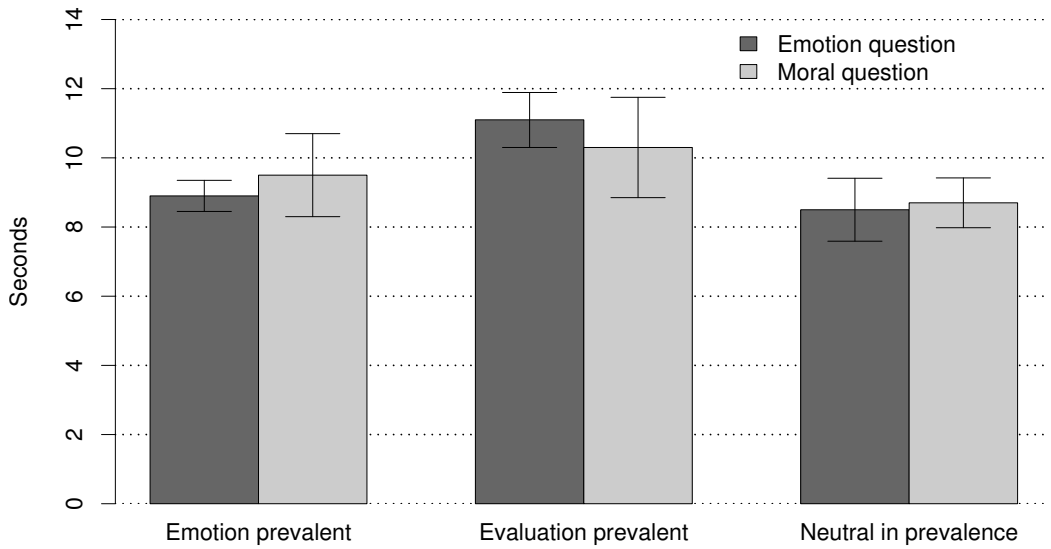

Figure 1: The latencies in Experiment 2 to response to the emotion and moral questions depending on whether in Experiment 1 the participants judged the scenarios to be emotion prevalent, evaluation prevalent, or neutral in prevalence.

3.2.2 Results and discussion

There was no reliable difference in the overall latencies to respond to emotion questions (9.15s), moral questions (9.91s), and consequential questions (9.38s; Friedman nonparametric analysis of variance, $\chi^2 = .25$, $p > .92$). Figure 1 presents the crucial latencies, those to respond to the emotion and moral questions depending on whether the scenarios were emotion prevalent, evaluation prevalent, or neutral in prevalence. As it shows, the predicted interaction occurred: the participants responded faster to emotion questions than to moral questions for the emotional prevalent scenarios, whereas they responded faster to moral questions than to emotion questions for the evaluation prevalent scenarios (Wilcoxon test, $z = 2.105$, $p < .02$). The positive scenarios that were emotion prevalent concerned actions of love, kindness, or friendship; and the negative scenarios concerned graphic violence or cannibalism. The positive scenarios that were evaluation prevalent concerned good actions with no striking emotional sequelae, such as the hiring of disabled individuals; and the negative scenarios concerned crimes without violence, such as perjury or bribery. The positive scenarios that were neutral in prevalence concerned care or cooperation; and the negative scenarios concerned sexual topics, and crimes against property.

For Humeans, emotions come first and cause moral evaluations. For moral grammarians, moral evaluations come first and trigger emotions "downstream". But, the reliable interaction in the latencies in the present experiment and the judgments in Experiment 1 tell a story that differs from both the Humean and grammatical accounts. The experimental results corroborate the principle of independent systems. Emotions sometimes precede evaluations, and evaluations sometimes precede emotions, and so it cannot be the case that one is always dependent on the other.

4 Evidence for prior conscious reasoning

The principle of deontic reasoning implies that naïve individuals often engage in conscious reasoning in order to reach a moral evaluation. As we mentioned earlier, not everyone accepts this view. Authors from Hume to Haidt have argued that conscious reasoning plays a subsidiary role, and no role whatsoever in your initial moral evaluations, which are driven solely by emotions or intuitions. In this section, we evaluate the findings contrary to this view and pertinent to a prior role of conscious reasoning at least on some occasions.

Piaget (1965/1932) carried out a series of informal studies on young children in order to test his theory of how they acquire the ability to distinguish between right and wrong. Both Kohlberg (1984) and he delineated a series of stages in moral development, but this topic is beyond the scope of the present paper. Our concern is solely with the evidence that Piaget reported from his dialogues with children. There were many such dialogues but we describe just a few typical examples. Consider these two contrasting scenarios:

> 1. Alfred meets a little friend of his who is very poor. This friend tells him that he has had no dinner that day because there was nothing to eat

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

> in his home. Then Alfred goes into a baker's shop, and as he has no money, he waits till the baker's back is turned and steals a roll. Then he runs out and gives the roll to his friend.
>
> 2. Henriette goes into a shop. She sees a pretty piece of ribbon on a table and thinks to herself that it would look very nice on her dress. So while the shop lady's back is turned (while the shop lady is not looking), she steals the ribbon and runs away.

The younger children in the study — those less than the age of ten — sometimes inferred the extent of moral transgressions in terms of their material consequences, and sometimes in terms of a protagonist's motives. The older children focused solely on motives. As an example of an evaluation based on material consequences, consider what one six-year old (S) said (Piaget, 1965, p. 131) in part of a dialogue with the experimenter (E):

> E. Must one of them [the two protagonists in the stories] be punished more than the other?
>
> S. Yes. The little boy stole the roll to give to his brother (sic). He must be punished more. Rolls cost more.

The child appears to be reasoning consciously:

> The little boy stole the roll.
>
> Therefore, he must be punished more than the girl who stole the ribbon because rolls cost more than ribbons.

Other children make analogous inferences based on the fact that the roll is bigger than the ribbon.

In contrast, a nine year-old took motive into account:

> E. Which of them is the naughtiest?
>
> S. The little girl took the ribbon for herself. The little boy took the roll too, but to give to his friend who had had no dinner.
>
> E. If you were the school teacher, which one would you punish most?
>
> S. The little girl.

This child also appears to be reasoning consciously, though relying on the unstated premise that those who do something wrong to benefit others are less culpable than those who do something wrong to benefit themselves.

Piaget reports many other dialogues based on scenarios illustrating contrasts of this sort, and the children appear to be reasoning consciously in order to reach moral evaluations. However, Humeans can argue that the children may have based their evaluations on a prior emotional response, and then used reasoning merely to try to convince the experimenter. One difficulty with this view is that it offers no account of how emotions lead children sometimes to focus on material consequences and sometimes to focus on intentions. Once again, it seems that we would need to invoke an emotional system capable of reasoning about these matters. Haidt's (2001) theory leads to an analogous problem if one asks how the children's conscious reasoning could influence the experimenter's intuitions. To follow a chain of conscious reasoning appears to depend on conscious reasoning. But, this process is precisely the one that is denied to the experimenter if conscious reasoning plays no part in eliciting moral evaluations. Nevertheless, Piaget's evidence is not decisive, because the children's reasoning may have been post hoc and not part of the process yielding their moral evaluations. In order to obviate this argument, we carried out a study in which adult participants thought aloud during the course of making moral evaluations.

4.1 Experiment 3

The aim of the experiment was to demonstrate that individuals do sometimes reason consciously in order to make a moral evaluation as opposed to reasoning only afterwards. A behavioral method for investigating this issue is to ask participants to think aloud as they are making a moral evaluation from information that forces them to reason. Introspections can be misleading evidence and yield only rationalizations (Nisbett & Wilson, 1977). But, when individuals think aloud as they reason, their protocols are a reliable guide to their sequences of thought (Ericsson & Simon, 1980) and to their strategies in reasoning: a program based on their reports can make the same inferences in the same way that they describe (Van der Henst, Yang, and Johnson-Laird, 2002).

Given a moral scenario to evaluate, individuals can make a snap moral evaluation and then engage in a subsequent process of conscious reasoning. Such protocols are consistent with an intuition preceding conscious reasoning. Another possibility is that individuals engage in a chain of conscious reasoning culminating in a moral evaluation. Such protocols are consistent with conscious reasoning determining the evaluation. Still another possibility is that individuals make a snap moral evaluation but immediately follow it up with a "because" clause explaining their reasons. Such protocols are ambiguous between the two previous cases. A skeptical Humean might argue that in all three cases what really comes first is an emotional reaction. But, a skeptical Kantian could counter that what really comes first is conscious reasoning. No

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

argument can rebut either sort of skeptic, but the issue then becomes untestable. Yet, the character of some protocols might strike all but the most dogmatic skeptics as good evidence for one sort of case or the other.

Each scenario described a single outcome, which was either moral or immoral, and two agents who played distinct causal roles: the action of one agent enabled the action of the other to cause the outcome. The participants had to judge which of the two agents was more praiseworthy for the moral outcomes, and which of the two was more blameworthy for the immoral outcomes. Previous studies have shown that naïve individuals distinguish between the two sorts of agents: enablers and causers (Frosch, Johnson-Laird & Cowley, 2007; Goldvarg & Johnson-Laird, 2001). Yet, the distinction between causes and enablers is so subtle that many philosophers, lawyers, and psychologists, have taken the view that it is, in Mill's term, capricious (Mill, 1843/1973). We spare readers the details of the current controversy about the meanings of causes and enablers, but the distinction between them should on occasion call for conscious reasoning. Consider the following scenario from the experiment:

> Barnett owned a gun store. He sold guns to everyone without checking IDs or whether the buyer had a criminal record. Martin came into the store intending to buy a weapon, and left with a handgun. He went home and fired it repeatedly. Later, his wife died from her wounds.

As it illustrates, the participants need to make a series of inferences to understand the causal sequence. They must infer that Barnett sold a handgun to Martin, because the scenario implies this proposition without stating it. They must similarly infer that Martin shot his wife. It follows that Barnett's action enabled Martin to shoot his wife. Previous experiments have shown that individuals are sensitive to this distinction in scenarios that state the relations more directly (e.g., Goldvarg & Johnson-Laird, 2001). Because causers are more responsible for outcomes than enablers (Frosch et al., 2007), the participants should infer that they are more praiseworthy for moral outcomes and more blameworthy for immoral outcomes.

4.1.1 Method

Eighteen volunteers in the Princeton University community (9 males and 9 females; mean age 22.6 years) took part in the experiment for payment. They were assigned at random to one of two independent groups: one group (10 participants) thought aloud as they tackled a scenario; the other control group (8 participants) did not. Participants in each group dealt with 6 scenarios (each of 50 words in length), presented in a different random order to each participant, but the same random order given to one participant in the think-aloud group was also given to one participant in the non-think-aloud group. The scenarios were based in part on those used by Frosch et al. (2007), three had moral outcomes, and three had immoral outcomes. We constructed two versions of each: one in which the enabler was described first, and one in which the causer was described first. Each participant tackled equal numbers of both sorts and both versions, which occurred equally often in the experiment as a whole. The following example illustrates a moral outcome with the causer described first:

> A visitor to the island had acute appendicitis. Despite a terrible storm with dangerous seas, Margie took her on a boat to the mainland. Tammy had always kept the boat ready for emergencies, with a full fuel tank, and a well-charged battery. The mainland surgeon operated to save the patient.

The participants were told that for each of a series of scenarios they had to decide which of two individuals was more morally praiseworthy or else more morally blameworthy. There was no time pressure. The participants in the think-aloud group were asked to think aloud in order to reach their decision.

4.1.2 Results and discussion

The participants in both groups tended to chose the causer rather than the enabler as more praiseworthy for good outcomes and more blameworthy for bad outcomes (83% of trials in the think-aloud group, and 83% of trials in the control group, Wilcoxon tests, $z = 2.72$, $p < .01$; $z = 2.34$, $p < .05$; notwithstanding a small but reliable tendency to choose the agent described second). This result suggests that the participants were reasoning about the contents of the scenarios, and that the task of thinking aloud did not have a major effect on evaluations. We classified the think-aloud protocols into three objective categories: those in which the participants stated a sequence of thoughts leading to a moral evaluation, and which were accordingly consistent with a consciously reasoned evaluation; those in which the participants made an immediate moral evaluation, and which were accordingly consistent with an initial intuition or emotional reaction; and those that were ambiguous because an immediate moral evaluation was followed at once with a "because" clause describing the reasons for the evaluation. Table 1 presents examples of the three sorts of protocol, and their overall percentages of occurrence. The participants as a whole showed a reliable tendency to make one or more reasoned responses more often than would occur by chance (Permutation test, $p < .05$). In addition, immediate decisions

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

Table 1: Three different sorts of think-aloud protocols with an example of each of them, and the percentages of their occurrence in Experiment 3. A *reasoned* protocol is one in which the participant consciously reasoned to reach a moral evaluation; an *immediate* protocol is one in which the participant made an immediate moral evaluation; and an *ambiguous* protocol is one that started with the moral evaluation but immediately appended a "because" clause reflecting a process of reasoning.

Response type	Example protocol	Example scenario	Percentages
Reasoned	Well this is a lot more... a lot more difficult to decide, because ultimately Martin is the one who made the decision to... commit the crime, but Barnett is the one who supplied the guns and... by law, I guess... Barnett would also... be at the same level of blame... that Martin is, but morally I feel that Barnett should... is definitely less blameworthy than Martin because he sold... he sells the guns and ultimately it's the decision of the consumer or whoever buys it how to use it. You know, you could use it for self defense, you can use it at a gun range... it's not really his responsibility to check... at least, from my point of view. Martin is definitely the one that's more blameworthy because he's the one that bought the gun and he's the one that committed the crime, and ultimately, if you think about it one way, Martin may not have actually needed the gun to commit the crime, I mean, he could've just gone home and just took a knife, and just stabbed... whoever he killed. So ultimately I feel that Martin's definitely the one that's more blameworthy.	Barnett owned a gun store. He sold guns to everyone without checking IDs or whether the buyer had a criminal record. Martin came into the store intending to buy a weapon, and left with a handgun. He went home and fired it repeatedly. Later, his wife died from her wounds.	35%
Immediate	Sid is more praiseworthy. Though the boss of the company allowed him to take a day off, he could do anything, but he decided to volunteer... and... do something good... so Sid is more praiseworthy, I think.	Zack was the boss of the small construction company for which Sid worked, and allowed him to take the day off without loss of pay. Sid joined a group of skilled volunteers building free houses for homeless people without accommodation. The volunteers built a new house for a poor person.	15%
Ambiguous	I would have to say that Peters is more morally blameworthy, because he's the one who... injured the man in this joke... Jones is just an honest mistake of leaving... he's careless, he left the elevator door open, whereas Peters's activity almost borders on... kinda maliciousness.	Peters, a young man who liked practical jokes, knowing that there was no elevator in the lift shaft, invited a visitor to step inside. The elevator in the apartment block was under repair, and Jones, the repairman, had carelessly left open the unguarded lift shaft. The visitor was badly injured.	50%

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

occurred less often than reasoned decisions or ambiguous decisions (Wilcoxon test over the participants, $z = 2.64$, $p < .01$). The participants were roughly divided between those who produced reasoned decisions on more than half the trials (4 out of 10) and the remainder who produced immediate or ambiguous decisions on more than half the trials. These results are consistent with a previous study suggesting that individuals differ in how they make moral evaluations (Moore et al., 2008), though most of our participants produced some consciously reasoned and some immediate evaluations.

5 Evidence for moral inconsistencies

The principle of moral inconsistency predicts the occurrence of moral conflicts that individuals may be unable to resolve. Hence, in tandem with the principle of deontic reasoning, it predicts not only that such conflicts should occur but also that individuals should be able to construct them for themselves. We carried out an experiment to test this prediction.

5.1 Experiment 4

5.1.1 Method

The experiment was advertised through a mailing list from Princeton's Psychology Department, and 21 participants (6 males, 15 females; mean age 23.9 years) carried it out on the World Wide Web. We chose three dilemmas (1–3) from Cushman et al. (2006), and devised three new dilemmas based in part on those in the literature. Each dilemma occurred in two versions. In the *contact* version, the agent made physical contact in order to carry out an action, and in the *non-contact* version the agent did not make physical contact in order to carry out an action. We summarize the two versions of each dilemma here:

1. Pushing a person in front of a trolley to kill him but to save five. Throwing a switch to push him.
2. Pushing a man out of the window to rescue five children. Swinging burning debris to push him.
3. Pushing a passenger overboard to save swimmers. Accelerating the boat so the passenger falls overboard.
4. Throwing the weakest person into the sea to save others. Ordering the crew to do so.
5. Operating on a woman to kill her but to save her triplets. Telling a surgeon to operate to do so.
6. Taking a girl off an artificial lung to save her brothers. Authorizing the doctors to do so.

Each participant dealt with only one version of a dilemma: three contact versions and three non-contact versions, which were presented in a different random order to each participant.

The participants carried out three tasks for each dilemma. First, they decided whether or not the action was permissible. Second, they modified the description of the dilemma so that they would switch their evaluation from permissible to impermissible, or vice versa. They were instructed not to eliminate the dilemma itself, e.g., the man still had to be pushed in front of the trolley. Third, they modified the description again to produce a version of the dilemma that was irresolvable for them.

5.1.2 Results and discussion

In the first task, the participants judged the actions to be impermissible for 62% of the contact dilemmas and 58% of the non-contact dilemmas (Mann-Whitney test: $z = .369$, $p > .7$), and so, contrary to previous studies, contact had no reliable effect on evaluations.

Table 2 below shows some typical examples of the results of the second task, in which the participants successfully modified the dilemmas so that they would switch their judgments: 15 out of the 21 participants were able to carry out this task for all six of the dilemmas, and all the participants except one were able to do so for over half of the dilemmas (Binomial probability for 20 out of 21 participants, $p << .001$). Many of their modifications resembled those that individuals make in thinking about counterfactual situations (Byrne, 2005). In order, say, to reverse their evaluation that an action was impermissible, they typically changed the status of the victim to one who volunteered to be sacrificed or to one who was a wicked individual. Conversely, to reverse their evaluation that an action was permissible, they changed the status of the victim to one who was a child or a family member.

Table 2 also shows some typical examples of the results of the third task, in which the participants successfully modified the dilemmas so that they would find them impossible to resolve: 14 out of the 21 participants were able to carry out this task for all six of the dilemmas, and again all the participants except for one were able to do so for over half of the dilemmas (Binomial probability, $p << .001$). Their main methods were again to modify the status of the victim (see the first two examples in Table 2), or the victim's desires (see the third example), or the efficacy of the action (see the fourth example). Only rarely did they alter a dilemma in a way that modified the status of the contact, or lack of contact, between the agent and the victim (in contrast to the salience of this variable in studies by Greene et al., 2004, and Hauser, 2006).

Hauser (2006) recognizes that moral dilemmas do occur. He writes: "Emotional conflict provides the telltale signature of a moral dilemma" (p. 223). The dilemmas themselves, he writes, "always represent a battle between

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

Table 2: Examples from Experiment 4 showing the participants' modifications of dilemmas a) to reverse their initial evaluations from permissible to impermissible, b) to reverse their initial evaluations from impermissible to permissible, and c) to make the dilemmas irresolvable.

Tasks	Examples of modifications
Revising from permissible to impermissible	"If the man is the sole provider for the five children, and without him, they will starve to death, then this act would be impermissible." (Fireman scenario)
Revising from impermissible to permissible	"If the brothers are highly respected scientists or leaders who are essential to the society, unplugging the girl's artificial lung might be permissible." (Comatose girl scenario)
Making irresolvable	"If the passenger were a close friend of his, like a brother to him, the son of a family that had taken him in after this parents died, someone he owed everything to and loved. and if the swimmers were his children." (Motorboat scenario) "If the man is a fellow fireman, it would be irresolvable [to push him through the window]." (Fireman scenario) "The sister will die soon, but has explicitly said that she does not want her kidneys removed from her body, because she fervently believes that her kidneys house her soul." (Comatose girl scenario) "The two brothers will also likely develop the same degenerative disease as their sister since both parents died from the same disease." (Comatose girl scenario)

different duties or obligations, and, classically, between an obligation to self versus other" (p. 194). This view is plausible; but presumably such dilemmas are ultimately resolvable on Hauser's account. Readers might suppose that the irresolvable conflicts in our experiment were not truly irresolvable, not even for those who devised them. Perhaps so, but the point is irrelevant to the purpose of the experiment. If a grammar decides moral issues for you, no irresolvable conflicts should occur, and you should not be able to devise them — any more than you could devise a string of words that was both grammatical and not grammatical in your natural language. And if to the contrary the grammatical theory allows that you should be able to construct dilemmas that you find irresolvable, the claim that moral evaluations are dependent on a *grammar* seems to be metaphorical rather than testable.

6 General discussion

Our aim has been to propose a theory of reasoning about moral propositions, and to corroborate its main predictions. The theory is based on earlier diverse accounts, which emphasize intuitions (Haidt, 2001, 2007), their innate basis (Hauser, 2006; Hauser et al., 2007), and their conflict with utilitarian reasons (Greene et al., 2001; Greene et al., 2004). But, the theory goes beyond each of these precursors. It is based on four fundamental principles:

1. Indefinability of moral propositions: No simple criterion exists to tell from a proposition alone whether or not it concerns morals as opposed to some other deontic matter, such as a convention, a game, or good manners.

2. Independent systems: Emotions and deontic evaluations are based on independent systems operating in parallel.

3. Deontic reasoning: all deontic evaluations, including those concerning morality, depend on inferences, either unconscious intuitions or conscious reasoning.

4. Moral inconsistency: the beliefs that are the basis of moral intuitions and conscious moral reasoning are neither complete nor consistent.

You recognize moral propositions, but you do not rely on any simple defining property, because, as we argued in laying out the first principle of the theory, no such criterion exists. Instead, you rely on your specific knowledge of your culture: you know what is and isn't a moral issue. You know, for instance, that in the West you pay interest on a mortgage, and that this matter is not normally a moral issue. Under the Sharia law of Islam, however, it is immoral to pay interest, and so banks make special provisions to finance the purchase of houses. What constitutes a moral issue is therefore often a matter of fact, and often a matter of the attitudes of the interested parties. It follows that reasoning about moral propositions is unlikely to depend on a special process, and the theory postulates

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

that it is merely normal deontic reasoning (Bucciarelli & Johnson-Laird, 2005). (We note in passing that there does not appear to be any special process of legal reasoning, either: it is merely normal reasoning about legal propositions.) The theory is consistent with a negative result from brain-imaging studies: "there is no specifically moral part of the brain" (Greene & Haidt, 2002, p. 522; pace Moll, de Oliveira-Souza, Bramati, & Grafman, 2002), and with Greene and Haidt's further conclusion that morality is probably not a "natural kind".

The principle of independent systems allows that emotions and deontic evaluations rely on systems that operate independently. Hence, some scenarios elicit an emotional response and then a moral evaluation – the emotion is prevalent, some elicit a moral evaluation and then an emotional response — the moral evaluation is prevalent, and some are neutral in that they elicit the two reactions at about the same time. We discovered in Experiment 1 that individuals tend to agree about which scenarios are in these different categories. When emotions were prevalent, the positive scenarios were about love, kindness, or friendship; and the negative scenarios were about violence or other horrific matters. When morality was prevalent, the positive scenarios were about good actions, such as helping disabled individuals; and the negative scenarios were about bribery, perjury, or other similar crimes without violence. The neutral scenarios were about cooperation or care in the positive cases, and about crimes against property or sexual topics in the negative cases.

The consensus was borne out in Experiment 2, which examined the latencies of the participants' responses to a question about the emotions evoked by the scenarios (does it make you feel good or bad?) and to a question about the morality of the scenarios (is it right or wrong?). Scenarios with prevalent emotions tended to elicit a faster emotional response, scenarios with prevalent evaluations tend to elicit a faster moral response, and neutral scenarios tend to elicit both sorts of response at about the same speed. The two systems — the emotional and the moral — are accordingly independent. Emotions in some cases can influence moral evaluations (Haidt, 2001), and moral evaluations in other cases can influence emotions (Hauser, 2006), and in still other cases, the two are concurrent (pace Haidt and Hauser). Indeed, some situations elicit a moral evaluation with little or no emotional overtones, e.g., you know it's wrong to steal a paper clip, and some situations elicit emotions with little or no moral overtones, e.g., you feel happy when you solve a difficult intellectual problem. Morals and emotions have no special interrelation any more than do problem solving and emotions.

The principle of deontic reasoning implies that all moral evaluations depend on inferences. Piaget (1965/1932) observed that young children are capable of arguments of the following sort:

> If someone does something wrong but to benefit someone else then they are not so naughty as someone who does something wrong for selfish reasons.
>
> The boy stole to benefit his friend.
>
> The girl stole for selfish reasons.
>
> Therefore, the boy wasn't so naughty as the girl.

But, his results did not show whether such reasoning leads to the moral evaluation. We accordingly carried out Experiment 3 in which the participants had to think aloud as they made a moral evaluation. All the participants were able to reason consciously in a way that led them to a moral evaluation. Consider the following scenario:

> Zack was the boss of the small construction company for which Sid worked, and allowed him to take the day off without loss of pay. Sid joined a group of skilled volunteers building free houses for homeless people without accommodation. The volunteers built a new house for a poor person.

Given this scenario and the question of which individual was more praiseworthy, one participant argued as follows: "It seems like... at first glance... both are morally praiseworthy, but if I had to choose one, I would choose Sid because he took more action and we don't know, technically, whether or not Zack knew what Sid was going to be doing, so I think that Sid is more morally praiseworthy for the actions he took on his day off."

Such a protocol is typical, and it strongly suggests that individuals do sometimes reason consciously in order to reach a moral conclusion. Of course, not all evaluations proceed in this way. Many appear to depend on intuitions based on unconscious premises. Is stealing wrong? Yes, you assent: you have a fundamental belief in the proposition. And why is it wrong? If you have a philosophical bent, you can try to construct an answer to the question or to find an underlying premise from which the belief follows. But, it could be that the proposition that stealing is wrong is axiomatic for you, and you are dumbfounded when you are asked to justify an axiom (Haidt, 2001).

Could it be that moral intuitions are based on the same unconscious cognitive evaluations that create emotions? Such a system would be parsimonious. But, it is unlikely for reasons that we have already adduced: moral intuitions may have no accompanying emotions; emotions may have no accompanying moral intuitions. Hence, the present theory posits three separate systems for emotions, intuitions, and conscious reasoning.

The principle of moral inconsistency postulates that the beliefs underlying your moral evaluations are neither complete nor consistent. It follows that situations can occur in which you are unable to offer a moral evaluation, or even to determine whether a moral issue is at stake. Is eating meat a moral issue? For many individuals it isn't: they eat meat because they like it, just as some vegetarians don't eat meat, because they don't like it. Other vegetarians, however, believe that eating meat is wrong. And some individuals may be unable to make up their mind: they eat meat, but wonder whether it is a moral issue and whether they should have a guilty conscience about it. Such uncertainties could not occur in a logical system of moral beliefs that was complete and consistent, just as a grammar gives a complete and consistent account of the set of sentences. A complete logical system would decide all moral issues without equivocation.

The principles of deontic reasoning and moral inconsistency imply that individuals should encounter, and indeed be able to construct, irresolvable dilemmas. Experiment 4 corroborated this prediction. The participants readily modified dilemmas to switch their judgments from permissible to impermissible, and vice versa. Similarly, they were able to modify them still further to construct dilemmas that they would find impossible to resolve. When naïve individuals construct a new version of a dilemma, whether to switch an evaluation or to make it impossible for them to resolve, they also appear to be engaged in conscious reasoning. It is most unlikely that naïve individuals could construct strings of sentences whose grammatical status was impossible for them to resolve. And so moral evaluation differs from sentencehood, and can hardly be based on a grammar in the usual sense of the term.

A common way to form an irresolvable dilemma is to make the victim, who is to be sacrificed to save other individuals, a relative or a close friend of the protagonist. Irresolvable dilemmas of this sort corroborate the principle of moral inconsistency. They also show that the utilitarian principle of the greatest good for society as a whole is not a binding normative principle in the deontic reasoning of daily life (see, e.g., Baron, 2008, Ch. 16; Sunstein, 2005). Friendship, special relationships, and even special individuals, can trump utilitarian head counts.

The present theory goes beyond other current accounts of moral reasoning in that it aims to dissolve any appeal to a special mechanism for moral reasoning. When you think about moral issues, you rely on the same independent mechanisms that underlie emotions and cognitions in deontic domains that have nothing to do with morality, such as games and manners. Your evaluations of the morality or immorality of actions depend, in turn, on unconscious intuitions or on conscious reasoning, but your beliefs do not always enable you to reach a clear decision about what is right and what is wrong, or even about whether the matter in hand is a moral issue.

References

Baron, J. (2008). *Thinking and deciding* 4th ed. New York: Cambridge University Press.

Bentham, J. (1996). *An introduction to the principles of morals and legislation*. (Ed. Burns, J. H., & Hart, H. L. A.). Oxford : Clarendon Press. (Originally published in 1789.)

Blair, R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, *57*, 1–29.

Borg, J. S., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience, 18*, 803–817.

Bucciarelli, M., & Johnson-Laird, P. N. (2005). Naïve deontics: a theory of meaning, representation, and reasoning. *Cognitive Psychology*, *50*, 159–193.

Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge, MA: MIT.

Cherubini, P., & Johnson-Laird, P. N. (2004). Does everyone love everyone? The psychology of iterative reasoning. *Thinking & Reasoning*, *10*, 31–53.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: The MIT Press.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgments: testing three principles of harm. *Psychological Science*, *17*, 1082–1089.

Davidson, P., Turiel, E., & Black, A. (1983). The effect of stimulus familiarity on the use of criteria and justifications in children's social reasoning. *British Journal of Developmental Psychology*, *1*, 49–65.

De Waal, F. (1996). *Good natured: the origins of right and wrong in humans and other animals.* Cambridge, MA: Harvard University Press.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, East Sussex: Psychology Press.

Fiddick, L., Spampinato, M. V., & Grafman, J. (2005). Social contracts and precautions activate different neurological systems: An fMRI investigation of deontic reasoning. *NeuroImage, 28*, 778–786.

Foot, P. (1967).The problem of abortion and the doctrine of the double effect. *Oxford Review*, *5*, 5–15.

Foot, P. (1972). Morality as a system of hypothetical imperatives. *The Philosophical Review*, *81* (3), 305–316.

https://doi.org/10.1017/S1930297500001479 Published online by Cambridge University Press

Frosch, C., Johnson-Laird, P. N., & Cowley, M. (2007). It's not my fault, your honor; I'm only the enabler. McNamara, D. S., & Trafton, J. G. (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, 1755. Nashville, Tennessee, USA.

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naïve causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*, 565–610.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, *6*, 517–523.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 2*, 389- 400.

Greene, J. D, Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.

Haidt, J. (2007). The new synthesis in moral psychology. *Science, vol. 316*, 998–1002.

Hare, R. M. (1981). *Moral thinking: Its levels, method and point*. Oxford: Clarendon Press.

Hauser, M. D. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: Harper Collins.

Hauser, M. D., Cushman, F., Young, L., Jin, K-X. R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language, 22*, 1–21.

Hill, T. E. (1992). *Dignity and practical reason in Kant's moral theory*. Ithaca: Cornell University Press.

Hopcroft, J. E., & Ulmann, J. D. (1979). *Formal languages and their relations to automata*. Reading, MA: Addison-Wesley.

Hume, D. (1978). *A treatise of human nature*. Second ed. Oxford: Oxford University Press. (Originally published 1739).

Hunt, E. B. (1999). What is a theory of thought? In Sternberg, R. J. (Ed.), *The nature of cognition* (pp. 3–49). Cambridge, MA: MIT Press.

Johnson-Laird, P. N. (2006). *How we reason*. Oxford: Oxford University Press.

Johnson-Laird, P. N., Legrenzi, P., Girotto, P., & Legrenzi, M. S. (2000). Illusions in reasoning about consistency. *Science*, *288*, 531–532.

Johnson-Laird, P. N., Mancini, F., & Gangemi, A. (2006). A hyper emotion theory of psychological illnesses. *Psychological Review*, *113*, 822–841.

Johnson-Laird, P. N., & Wason, P. C., Eds. (1977). *Thinking*. Cambridge: Cambridge University Press.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In Holyoak, K. J., & Morrison, R. G. (Eds.) *The Cambridge handbook of thinking and reasoning*. (pp. 267–293). Cambridge: Cambridge University Press.

Kant, I. (1959). *Foundations of the metaphysics of morals,* and *What is enlightenment?* Trans. By L. W. Beck. Indianapolis: Bobbs-Merrill. (Originally published 1785)

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature, vol. 446*, 908–911.

Kohlberg, L. (1984). *The psychology of moral development: the nature and validity of moral stages*. San Francisco: Harper & Row.

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Sciences*, *114*, 143–152.

Mill, J. S. (1973). *A system of logic ratiocinative and inductive : Being a connected view of the principles of evidence and the methods of scientific investigation*. Toronto: University of Toronto Press, Routledge & Kegan Paul. (Originally published 1843.)

Mill, J. S. (1998). *Utilitarianism*. Oxford: Oxford University Press. (Originally published 1863.)

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press. Cambridge: Cambridge University Press.

Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *Neuroimage, 16*, 696–703.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, in press.

Nichols, S. (2002). Norms with feeling: towards a psychological account of moral judgment. *Cognition*, *84*, 221–236.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Oatley, K., & Johnson-Laird, P. N. (1996). The communicative theory of emotions: Empirical tests, mental models, and implications for social interaction. In Martin, L. L., & Tesser, A. (Eds.) *Striving and feeling: Interactions among goals, affect, and self-regulation*. (pp. 363–393). Mahwah, NJ: Erlbaum.

Piaget, J. (1965). *The moral judgment of the child*. New York: Free Press. (Originally published in 1932.)

Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility.

Journal of Experimental Social Psychology, *39*, 653–660.

Reitman, W. R. (1965). *Cognition and thought*. New York: Wiley.

Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, *15*, 165–184.

Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century*. New York: W.H. Freeman.

Sloman, S. A. (1996).The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Suber, P. (1990). *The paradox of self-amendment: A study of logic, law, omnipotence, and chance*. Berlin: Peter Lang.

Sunstein, C. R. (2005). Moral heuristics. *Behavioral and Brain Sciences*, *28*, 531–573.

Van der Henst, J-B., Yang, Y., & Johnson-Laird, P. N. (2002). Strategies in sentential reasoning. *Cognitive Science*, *26*, 425–468.

Wainryb, C., & Turiel, E (1993). Conceptual and informational features in moral decision making. *Educational Psychologist*, 28, 205–218.

Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science, 18*(3), 247–253.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*, 780–784.