

# Chapter 8

## Moral Brains – Possibilities and Limits of the Neuroscience of Ethics

Kristin Prehn and Hauke R. Heekeren

### 8.1 Introduction

A “neuroscience of ethics” focuses on the question of what can be learned about morality or moral standards (standards that an individual or a group has about what is right and wrong or good and evil) through a growing understanding of how the human brain works.

Empirical research in the field of morality comprises two main questions: How people generally distinguish “right” from “wrong,” and how people behave in a morally appropriate way, for instance, resisting the temptation to do wrong. Both questions, as well as the question regarding an interrelation of moral judgment and behavior, have been of recurring interest in many disciplines including philosophy, arts, religion, or law studies. In the field of psychology (the science aiming to understand and predict human behavior), in particular, a variety of theories and models have been developed to explain moral judgment. Interestingly, in most psychological approaches moral judgment is regarded as a precondition for moral behavior and defined as the evaluation of one’s own or someone else’s behavior with respect to social norms and values considered to be virtuous by a culture or subculture, such as not stealing or being an honest citizen (definition adapted from Haidt 2001: 817).

In recent years, advances in cognitive neuroscience have provided new technologies, such as functional magnetic resonance imaging (fMRI), that make it possible to investigate the neural substrates of moral judgment and behavior (to “localize the moral brain”). Since the advent of these methods, the question of how and where morality is located in the human brain has triggered much research. Research studies question which cognitive processes are involved, to what extent these processes are open to conscious deliberation, and whether human moral behavior is a product of education or a result of an innate mechanism activated during childhood.

---

K. Prehn (✉) • H.R. Heekeren  
Cluster of Excellence “Languages of Emotion”, Freie Universität Berlin, Berlin, Germany  
e-mail: kristin.prehn@fu-berlin.de

In particular, the question of whether moral judgments are caused by emotional or cognitive processes and whether emotional responses make moral judgments better or worse has caused much controversy and debate.

In the following chapter, we will, first, give a brief overview of traditional and recent psychological models of moral judgment and behavior (Sect. 8.2). Second, we will introduce the neuroscientific approach and the methods applied to the study of the “moral brain,” including the examination of brain damaged patients, neuroimaging, and neurostimulation (Sect. 8.3). Then, we will present main lines of research and give a critical overview of some studies aiming to disentangle domain-specific and -general processes involved in moral judgment and behavior and, finally, present our own empirical findings based on a neuroimaging study investigating the influence of individual differences in moral judgment competence (according to the Dual Aspect Theory by Georg Lind; Sect. 8.4).

## 8.2 Psychological Models on Moral Judgment and Behavior

### 8.2.1 *Moral Reasoning Investigated from a Cognitive-Developmental Perspective*

Psychological research on morality has long been dominated by a cognitive-developmental approach, investigating the maturation of moral reasoning and its underlying moral orientations and principles as a precondition for moral behavior (Piaget 1965; Kohlberg 1969).

To investigate the maturation of moral reasoning, Lawrence Kohlberg presented children and adolescent participants with moral dilemmas and asked them to argue why it could be justified to choose a certain action. In one of his best known dilemmas (“Heinz dilemma”), for instance, a man named Heinz has to decide if he should break into a drugstore to steal a medicine that would save the life of his dying wife. Based on how children and adolescents argued, Kohlberg established his much cited six-stage model (three levels including two stages at each level) of cognitive development of moral reasoning. This model proposes that humans progress through six stages as their cognitive abilities mature. During this development, people acquire a more sophisticated understanding of social relationships and, in particular, come to see situations not only from their own perspective but also from the perspectives of all other people involved in the conflict.

According to Kohlberg, at the pre-conventional level, young children think a behavior is right when an authority says it is. Doing the right thing means obeying an authority and avoiding punishment (stage 1 = obedience and punishment orientation). At stage 2 (= self-interest and exchange orientation), children see that there can be different sides to an issue and each person is free to pursue his or her own interests. Additionally, children understand that it is often useful to do someone else a favor. Later, at the conventional level, young people think of themselves

as members of their society with its values, norms, and expectations. At stage 3 (= interpersonal accord and conformity orientation), they aim to be a “good boy or girl,” which basically means being helpful to other people who are close to them. At stage 4 (= authority and social order maintaining orientation), the concern shifts toward obeying the laws to maintain society as a whole. At the post-conventional level, people start to think about the principles and values that constitute a good society. At stage 5 (= social contract orientation), laws are regarded as social contracts rather than rigid dictums. Those laws that do not promote the general welfare should be changed when necessary to meet the greatest good for the largest number of people (e.g., by democratic majority decisions). Finally at stage 6 (= universal ethical principles), moral reasoning is thought to be based on abstract reasoning using universal ethical principles of justice and of the reciprocity and equality of human rights with respect for the dignity of human beings as individuals (Kohlberg 1969).

For our purposes, in order to explore which processes are involved in moral judgment and behavior, the relevance of the cognitive-developmental theory could be seen as the idea that morality does not only rely on the acquisition of social norms and values held to be virtuous in a community (i.e., the acquisition of social knowledge), but also on the way individuals understand and think about social situations. Following Kohlberg, how people think about social situations qualitatively changes as result of an active interaction of the individual with his or her social environment.

It is also noteworthy that Kohlberg defined morality from his developmental perspective in terms of an ability, as “the capacity to make decisions and judgments which are moral” (i.e., based on internal moral principles and to act in accordance with such judgments; Kohlberg 1964: 425). Based on this notion of morality as an ability, Georg Lind in a current theoretical approach (Dual Aspect Theory) defines morality as consisting of two inseparable, yet distinguishable aspects: (a) a person’s moral orientations and principles and (b) a person’s competence to act accordingly. Following the Dual Aspect Theory, moral judgment competence is the ability to apply moral orientations and principles in a consistent and differentiated manner in varying social situations. Thus, social norms and values (represented in the Dual Aspect Theory as affect-laden moral orientations and principles) are linked with everyday behavior and decision making by means of “moral judgment competence.” Moral judgment competence represents a cognitive component, regarded as an important condition for living together in a democracy. Moral judgment competence can be trained by interventions such as the Konstanz Method of Dilemma Discussion (KMDD), developed by Lind to improve pro-social behavior, learning and decision-making skills, affect regulation, and the prevention of antisocial behavior (Lind 2008).

Although the cognitive-developmental theory has strongly influenced the discourse about morality and the subsequent research on moral education, there is also some criticism. It has been criticized, for instance, that Kohlberg investigated only post-hoc justifications for moral judgments that already had occurred, rather than actual reasoning processes leading to moral judgments (see Sect. 8.2.2; Haidt 2001).

Moreover, the assumption of a universal and invariant sequence of developmental stages has been doubted (Snarey 1985). Another point of criticism is that Kohlberg's theory emphasizes justice to the exclusion of other values. Carol Gilligan, in particular, argues that Kohlberg's theory is mainly based on empirical research in male participants and thus does not adequately describe the concerns of women. Therefore, she developed an alternative theory of moral reasoning that is not based on justice but on the ethics of caring (Gilligan 1977; Gilligan and Attanucci 1988; for recent neuroscientific studies investigating differences between justice and care ethics, see Robertson et al. 2007; Cáceda et al. 2011).

### ***8.2.2 The Role of Emotion and Intuition in Moral Judgment and Behavior***

More recent theories and models question the importance of rational reasoning processes for morality and emphasize the impact of intuitive feelings and automatic emotional responses.

James Blair (1995), for instance, suggested that humans (similar to other animals) possess a mechanism which, when activated by the communication of distress, such as sad facial expressions or tears, mediates the suppression of aggression (a so-called violence inhibition mechanism, VIM). He claimed that the VIM is a precondition for the development of (1) moral emotions such as sympathy, guilt, and remorse, (2) non-violent behavior, and (3) the moral/conventional distinction during childhood. This latter distinction between moral and conventional transgressions found in the judgments of children and adults marks the ability to differentiate cases where harm is caused to a person (= moral transgressions) from cases where only socio-conventional norms are violated (= conventional transgressions) without necessarily causing harm (e.g., spitting in a glass of wine at a dinner party; see also Turiel 1983; Nichols 2002). Specifically, Blair proposes that a lack of the VIM would explain the core symptoms of psychopathy and his empirical study could demonstrate that psychopaths—which according to the diagnostic criteria show an early onset of extremely aggressive and violent behavior and a lack of moral emotions like sympathy, guilt, and remorse—also fail to differentiate between moral and conventional transgressions in contrast to healthy controls (Blair 1995).

The social intuitionist model by Jonathan Haidt (2001) is another theory suggesting that fast and automatic intuitions like gut feelings or aesthetic judgments are the primary source of moral judgments, whereas rational arguments as obtained in Kohlberg's interviews are only used to construct post hoc justifications for judgments that have already occurred. "Moral intuition" is defined as the sudden appearance of a moral judgment in consciousness including a strong affective valence (good vs. bad, like vs. dislike). This would mean that reasoning is less relevant to moral judgment and behavior than Kohlberg's theory suggests and implies that people often make judgments without weighing concerns such as fairness, law, human rights, or abstract ethical values. Haidt illustrates the alleged minor role of rational

reasoning in moral judgment provocatively as the “rational tail of the emotional dog” and provides some striking examples of “moral dumbfounding” in which participants were unable to generate adequate reasons for an intuitively given moral judgment. When presented with the case of consensual sex between adult siblings, for instance, almost everyone reports a strong emotional response and a feeling that it is wrong, even though he or she cannot articulate reasons for this opinion. Further highlighting the role of emotion, Haidt (2003) suggests some useful distinctions, sorting moral emotions (i.e., emotions in response to moral violations that motivate moral judgments and behavior) into other-condemning emotions (contempt, anger, and disgust), self-conscious emotions (shame, embarrassment, and guilt), the other-suffering family (sympathy and compassion), and the other-praising family (gratitude, awe, and elevation).

Similarly, the universal moral grammar theory proposes that the human mind is endowed with an innate moral grammar consisting of a domain-specific, complex set of rules, concepts, and principles that guide human social behavior in a community (by using concepts and models analogous to those used in the study of language; e.g., Hauser 2006b; Mikhail 2007). There is evidence, in fact, that people consistently judge harm caused by action as morally worse than the same harm caused by omission (action principle). Harm intended as means to an end is also judged as morally worse than the same harm foreseen as a side effect of reaching a goal (intention principle). Using physical contact to cause harm to a victim, moreover, is judged as morally worse than causing equivalent harm to a victim without using physical contact (contact principle).

Although there seems to be much evidence supporting the role of moral intuitions and principles in moral judgment and behavior (e.g., Haidt et al. 1993), other researchers qualify the strong assertions of the social intuitionist model and the universal grammar theory by pointing out that immediate intuitions and moral principles can also be informed and shaped by conscious reasoning (e.g., Pizarro and Bloom 2003; Takezawa et al. 2006). At least this is the case when participants have enough time to deliberate thoroughly (Suter and Hertwig 2011). Some principles, however (such as the intention principle with its distinction between intended and foreseen consequences) appear to be inaccessible to conscious reflection (see Cushman et al. 2006).

Incorporating psychological, developmental, and evolutionary perspectives, Haidt and Joseph recently proposed the moral foundation theory (MFT, Haidt and Joseph 2007). The MFT proposes that morality (perceived as a broad concept going beyond questions of harm and fairness) is built upon five innate and universally available “foundations” that have been selected through human evolution and are shaped during a person’s individual development. The five foundations are: harm/care, fairness/reciprocity, in-group/loyalty, authority/respect, and purity/sanctity. Empirical studies aimed at verifying the MFT, for instance, showed that people with different cultural and political backgrounds (e.g., liberals vs. conservatives) differ in the degree to which they endorse each of the five moral systems (Graham et al. 2009). Glenn et al. (2009), moreover, found that higher scores in a measure of psychopathy predicted lower scores on the harm/care and fairness/reciprocity

subscales of a measure of the moral foundations, but showed no relationship with authority/respect, and only small correlations with in-group/loyalty and purity/sanctity. On a measure of “willingness to violate moral standards for money,” psychopathy scores predicted greater willingness to violate moral concerns of any type. While the moral foundations approach enjoys a growing popularity (see e.g., [www.moralfoundations.org](http://www.moralfoundations.org)), it must be stated that value and validity of this theory have already been put into question (see Suhler and Churchland 2011).

According to all three theoretical approaches highlighting the role of emotion (social intuitionist model, moral grammar and moral foundations theory), human morality relies at least to some degree on intuitive feelings and mechanisms, which are in part thought to be innate. Furthermore, it is stated that humans often have no conscious understanding of why they feel what they feel. The love felt toward one’s own children and the anger felt toward someone who cheated on us can thus be considered as an adaptive mechanism of selective advantage that was shaped over the course of evolution (for further evolutionary considerations, see Prehn and Heekeren 2009).

### **8.3 The Neuroscientific Approach Investigating Morality**

#### ***8.3.1 Lesion Studies Provide First Evidence of a Neurobiological Basis of Morality***

A first hint indicating that morality (i.e., moral judgment and behavior) might have a neurobiological basis stems from the classic case of Phineas Gage, a railroad worker whose ventromedial prefrontal cortex (VMPFC) was damaged in an accidental explosion (Harlow 1848; Damasio et al. 1994). After his recovery, he showed preserved basic cognitive abilities and social knowledge (as indexed by IQ-tests and other measures) but an irresponsible and inappropriate social behavior, impaired decision making in everyday life, and a limited ability to experience emotions (a so-called “acquired sociopathy,” Damasio et al. 1994).

More recent lesion studies report that damage to the prefrontal cortex (specifically, its ventromedial and orbitofrontal portions) leads to deficits in moral emotions, social behavior, and decision making (Saver and Damasio 1991; Barrash et al. 2000; Ciaramelli et al. 2007; Koenigs and Tranel 2007; Koenigs et al. 2007; Moretto et al. 2010; Thomas et al. 2011; Young et al. 2010b). For instance, it has been demonstrated that patients with lesions in the orbitofrontal cortex (OFC) display a defective ability with regard to anticipating negative consequences of one’s choices during a gambling task, and they also do not experience regret afterwards (Camille et al. 2004). Notably, the age at which a brain injury occurred has been found to affect the degree and nature of the deficits. Anderson et al. (1999) showed that lesions in the VMPFC and OFC acquired in early childhood not only lead to impaired social and moral behavior but also seem to prevent the acquisition of factual knowledge about

the accepted standards of moral behavior in general (see also Eslinger and Biddle 2000). In sum, lesion studies provide evidence that at least some of the processes involved in moral judgment and behavior are dissociable (e.g., the distinction between acquisition and application of social rules mentioned above or identifying specific subcomponents such as the ability to anticipate punishment or to experience moral emotions).

Notably, lesion case studies have contributed significantly to theory evolution in the field of moral cognition. Antonio Damasio, for instance, posited his “somatic marker hypothesis” based on his observations of patients with lesions of the VMPFC (Damasio et al. 1994; Damasio 1996). The somatic marker hypothesis suggests that emotional responses involving body function changes (labeled as “somatic markers”), such as an increase in heart rate or skin conductance, become associated over time with reward or punishment. After the repeated experience of certain bodily changes as response to the outcome of a certain action, such as a bad feeling when caught red-handed, the brain areas that monitor these bodily changes begin to respond whenever a similar situation with similar behavioral options arises. According to this theory, somatic markers are integrated in the VMPFC with other knowledge and planning functions and, thus, bias real life decision making in the future, especially in very complex situations with a high degree of uncertainty and ambiguity.

### ***8.3.2 Neuroimaging Reveals a Distributed Functional Network Involved in Moral Cognition***

It is important to keep in mind that lesion studies usually rely on a very limited number of cases with mostly very large and heterogeneous lesions. They give important hints (e.g., about single and double dissociations of cognitive processes) but cannot really reveal how the process of behaving appropriately or making moral judgments and ethical choices is organized in an intact human brain. In an attempt to overcome this limitation, cognitive neuroscientists have taken great advantage of the development of neuroimaging methods like functional magnetic resonance imaging (fMRI) which enables researchers to measure brain activity of healthy participants during a specific moral task such as judging a described behavior as being good or bad.

fMRI was first used in humans in 1991 (Belliveau et al. 1991). In its most popular variant, it measures cerebral changes of local hemoglobin oxygenation in response to a certain task (see Logothetis 2008 for a review). The method is based on the fact that the execution of a task leads to increased neuronal activity in the brain regions preoccupied with its processing. Increased neuronal activity is accompanied by a depolarization of neuron membrane potentials. Maintaining and re-establishing these potentials in groups of neurons requires an increased supply of energy and oxygen. This, in turn, leads to an increase in blood flow and blood volume in the capillaries of the activated brain tissue (commonly referred to as

“neurovascular coupling”) resulting in both an increase of oxygenated hemoglobin, which overcompensates the actual supply of oxygen, and a concomitant decrease in deoxyhemoglobin concentration in this brain region. The changes of the local blood flow and blood volume as well as the relative change of deoxyhemoglobin in the blood concentration determine the so-called blood-oxygen level dependent signal (BOLD-signal) which can be detected due to the paramagnetic properties of deoxyhemoglobin by an MRI scanner with a powerful magnet (typically, 1.5 or 3.0 T). Although undoubtedly revolutionary for the study of mental phenomena, neuroimaging has some specifics that should be kept in mind when discussing its results.

First of all, it is important to know that during the performance of a task (e.g., when making a moral judgment or, in principle, at any time of wakeful activity) many if not all parts of the brain are activated to some degree. To identify brain regions that are specifically related to morality, most researchers are using “subtraction logic” in their experimental designs. Subtraction logic was pioneered by the Dutch physiologist Franciscus Cornelius Donders in reaction time experiments (see Donders 1969, translation of: *Die Schnelligkeit psychischer Prozesse*, first published in 1868, *Archiv für Anatomie und Physiologie*, 8, 657–681). The concept is based on the assumption of “pure insertion,” which means that one cognitive process can be added to a pre-existing set of cognitive processes without affecting them and asserts that there are no interactions among the different components of a task. Although this assumption has not been validated in any physiological sense (Friston et al. 1996), it is applied in almost all fMRI studies mentioned in this chapter. In one of our recent studies (see Sect. 8.4.3; Prehn et al. 2008), for example, we compared neural activity during a moral judgment task with activity during a grammatical judgment task. The grammatical judgment task was designed to share almost all processes with the moral judgment task except the moral component: During both tasks, participants had to read sentences on a screen, to decide whether the actions described were “correct” or not (morally or grammatically), and then to respond with a button press. The grammatical judgment task, thus, controls for visual input, language processing, decision making, and motor output. In other words, colorful pictures of brains “lighting up” are actually artifacts of statistical analysis and selective presentation. They show those brain regions where a statistically significant level of increase or decrease in BOLD signal occurred during a task relative to a control state. In addition, results are mostly based on some kind of accumulation over a sample of only 20–30 subjects.

Following from the need to apply subtraction logic, data on the neural correlates of mental phenomena can only be as good as the underlying tasks and experimental paradigms. Experimental tasks have to be carefully designed so that they specifically activate the cognitive functions of interest and avoid the presence of other “confounding” factors that could possibly serve as an alternative explanation of the observed effects. The need to control for confounding factors in an experimental design prompts researchers to sometimes strip away real-life contexts and thereby undercut the “ecological validity” of a study.

fMRI studies investigating moral judgment and behavior have employed very different types of tasks and stimuli. As the study of morality has been traditionally



based in the domain of philosophy, many investigators have used complex moral dilemmas similar to those discussed by contemporary moral philosophers (e.g., Greene et al. 2001, 2004; Young et al. 2007; see Sect. 8.4). Other types of stimuli have been short sentences containing social norm violations (e.g., Heekeren et al. 2003, 2005; Prehn et al. 2008) as well as pictures or picture sequences with moral content (e.g., Bahnemann et al. 2010; Moll et al. 2002b). Some researchers invented innovative paradigms for the study of honest or dishonest “cheating” behavior (Greene and Paxton 2009), the making of charitable donations (Moll et al. 2006), or acts of reactive aggression and punishment (Buckholz et al. 2008; Lotze et al. 2007). To study cooperative, altruistic, or self-interested behavior, economic decision-making tasks such as the Ultimatum Game or Reciprocal Trust Game have been used, during which two participants interact with each other via a computer interface and decide how to divide a given sum of money (Rilling et al. 2002; Sanfey et al. 2003; de Quervain et al. 2004; Spitzer et al. 2007). To investigate social interaction processes in more detail, a very interesting method is hyperscanning, by which two or even multiple subjects, each lying in a separate MRI scanner, can interact with one another while their brains are simultaneously scanned. Hyperscanning permits the study of brain responses that underlie processes during social interactions (for examples of scanning two participants to compute “between brain correlations,” see Montague et al. 2002; Krueger et al. 2007).

Using these different tasks to investigate the variety of moral phenomena, neuroimaging studies have been remarkably consistent in revealing a functional network of brain regions involved. This network includes prefrontal brain regions, such as the ventromedial prefrontal cortex (VMPFC), the orbitofrontal cortex (OFC), and the dorsolateral prefrontal cortex (DLPFC), the temporal poles, the amygdala, the posterior cingulate cortex (PCC), the posterior superior temporal sulcus (pSTS), as well as the temporo-parietal junction (TPJ; for reviews, see Greene and Haidt 2002; Moll et al. 2003, 2005b, 2008a; Casebeer 2003; Lieberman 2007; Young and Dungan 2011).

The relative activation of a particular brain region during one experimental condition compared with another, however, does not tell us that much by itself. This information is only “spots on brains” until it is related to a hypothesis and to the developing picture of cognitive localization and integration in the brain. Complex tasks, such as judging whether a presented behavior is wrong in regard to moral conventions, comprise numerous cognitive and affective processes even when compared with a perfectly designed control task. These processes are represented by a distributed network of brain regions. Therefore, we cannot expect morality to be located in a specific and distinct brain area (“a moral center”). On top of that, different tasks often show highly overlapping neural networks. Processes thought to be different (such as emotion and cognition) are not necessarily subserved by separate and independent circuits (cf. Pessoa 2008). To be able to interpret a certain pattern of brain activity as a response to a specific task, one therefore needs very clear hypotheses about the involved mental processes. Such hypotheses can be derived from psychological theories or assumptions about the underlying neuronal mechanisms, for instance, resulting from lesion data or electrophysiological studies in monkeys and apes. A particularly nice way of linking imaging data with a targeted

function involves establishing some kind of intensity measure from the behavioral data (such as response times and post-hoc ratings), or the assessment of individual differences in personality traits, attitudes, and abilities. These measures can be used to demonstrate a corresponding change of intensity in the imaging data (on the question how to infer mental processes from imaging data, see Henson 2006; Poldrack 2006).

### ***8.3.3 Neurostimulation Methods as an Attempt to Modulate Activity in the “Moral Brain”***

In addition to the limitations already mentioned, it is important to understand that neuroimaging only allows us to see the changes in brain activity that are correlated with an experimental condition. Showing that one brain region is activated during a task does not show that this brain region is actually used or even necessary for the task. Neurostimulation methods like transcranial magnetic and direct current stimulation go beyond this correlational approach and can be used to demonstrate causality. If a subject performs worse on a task after a specific brain region was knocked out by an induced electric current for the time of task processing (also known as virtual lesion approach), this is much stronger evidence that this region is actually involved in performing the task.

Transcranial magnetic stimulation (TMS) is a non-invasive technique that is used to induce weak electric currents in the cortical tissue of the brain. TMS was first introduced for the investigation of the motor cortex by Barker and colleagues in 1985, who demonstrated that a magnetic pulse caused by a coil placed over the motor cortex produces an action potential (nerve impulse) which is transmitted from the cortex to the spinal cord and leads to a subsequent muscle contraction of the contralateral hand (Barker et al. 1985). The repetitive application of magnetic pulses (called repetitive transcranial magnetic stimulation=rTMS) in healthy participants is nowadays used to study a variety of cerebral functions either causing excitation or inhibition of neural activity (high-frequency rTMS leads to neuronal depolarization and enhanced neural firing, whereas low-frequency rTMS has been found to disrupt neural activity in a cortical area; see Guse et al. 2010).

As we will show in greater detail in the next sections, rTMS has been successfully used in the study of morality. For instance, it has been argued that the capacity to infer the actor's intentions and beliefs is central to moral judgment, which is associated with activity in the TPJ. Young et al. (2010a) showed that a disruption of neural activity in the right TPJ alters moral judgments insofar that participants in the TMS condition judged cases as less morally blameworthy when actors intended but failed to do harm than cases in which harm was caused accidentally. By using moral dilemmas that induce a conflict between emotion and reason (see Sect. 8.4.1), it was, moreover, found that a disruption of neural activity in the right DLPFC alters moral judgment and leads to an increase of utilitarian responses (Tassy et al. 2012).

Transcranial direct current stimulation (tDCS) is another non-invasive tool for modulating cortical excitability. Although already developed in the 1960s, this method has only recently been studied more extensively, with the advent of other brain activation techniques such as TMS (also with regard to its potential clinical application). TDCS protocols basically involve the application of two surface electrodes on the scalp of the participant, one serving as the anode and the other serving as the cathode. A 1–2 mA direct current then is applied for up to 20 min between the two electrodes, flows from the anode to the cathode, and leads to increases or decreases in cortical excitability dependent on the direction of the current. Anodal tDCS results in depolarization of the neurons underneath the electrode, hence causing an excitatory effect, whereas cathodal tDCS results in hyperpolarization and thus inhibition of cortical neurons (Been et al. 2007). In contrast to TMS, tDCS does not directly elicit action potentials (by means of suprathreshold resting membrane potential change) but renders neuronal populations more or less ready to fire in response to additional inputs. In other words, tDCS changes the likelihood that an incoming action potential will result in postsynaptic firing.

A number of studies has shown that tDCS applied to the prefrontal cortex has effects on cognition and mood. With regard to morality, it has been found that anodal stimulation over the right DLPFC (which results in an upregulation of neural activity) reduces risk-taking during decision making (Fecteau et al. 2007a, b). In contrast, an inhibition of the anterior prefrontal cortex through cathodal stimulation improves deceptive behavior; that is, it leads to better lying skills, reduced skin conductance responses, and feelings of guilt (Karim et al. 2010).

When following established safety protocols, both neurostimulation methods are safe and do not cause any side effects, apart from mild headache, discomfort because of unintended stimulation of nerves and muscles on the head, or itching underneath the electrodes (see Nitsche et al. 2003; Rossi et al. 2009). Although magnetic pulses and direct currents can only be administered on the surface of the cerebral cortex (only approximately 2 cm below the scalp), neurostimulation is not limited to cortical regions. Since the brain is an interconnected system, neuromodulation can also occur at distant but interconnected regions, such as deep brain structures like the amygdala (via its connections to the prefrontal cortex). In sum, neurostimulation methods offer an interesting perspective not only to the study of morality but also to a potential modulation of moral judgment and behavior (for instance, in therapeutic settings).

#### **8.4 Studies Investigating the Role of Domain-Specific and General Capacities Contributing to Moral Judgment and Behavior**

In the last decade, an increasing number of neuroimaging studies has been conducted to investigate the neural correlates of moral judgment. Some studies have focused on the neural correlates of moral judgment in general and in comparison to other non-moral (e.g., Moll et al. 2001, 2002a; Heekeren et al. 2003) or aesthetic

judgments (Tsukiura and Cabeza 2010). Others investigated the neural correlates of specific moral emotions (guilt, shame, regret and moral disgust or indignation; e.g. Coricelli et al. 2005; Moll et al. 2005a; Wagner et al. 2011). Moreover, studies have focused on the evaluation of one's own or other agents' actions and whether it matters if harm was caused intentionally or accidentally (e.g., Berthoz et al. 2002, 2006; Schaich Borg et al. 2006; Young et al. 2007; Young and Saxe 2008), on the influence of bodily harm on neural correlates of moral decision making (Heekeren et al. 2005), on the regulation of emotional responses (Harenski and Hamann 2006), and the impact of audience on moral judgments (Finger et al. 2006). Recent work was also dedicated to a differentiation of moral intuition and moral reasoning (Harenski et al. 2010a).

Neuropsychiatrists dealing with antisocial individuals and the biological foundations of criminal behavior have also contributed to the study of morality and have applied structural and functional MRI to investigate the "immoral brain" in clinical populations with difficulties in moral judgment and behavior, such as psychopaths (de Oliveira-Souza et al. 2008; Harenski et al. 2009, 2010b; Harenski and Kiehl 2010; Prehn et al. 2013). Kent Kiehl, in particular, has contributed enormously to the field by traveling with a mobile MRI scanner mounted on a truck and investigating more than 1,000 prison inmates. He linked emotional hypo-reactivity found in psychopaths to a dysfunctional paralimbic system including anterior and posterior cingulate, insula, OFC, amygdala, parahippocampal gyrus and superior temporal gyrus (Kiehl 2006; for reviews, see also Raine and Yang 2006; Blair 2008; Glenn and Raine 2008).

Below, we will look at three research foci, dedicated to the question of how and where moral judgment is processed in the human brain: (1) the relationship of emotional and cognitive subsystems contributing to moral judgment and behavior, (2) the role of social cognitive processes and mental state reasoning, and (3) the influence of individual differences (see also the review by Young and Dungan 2011).

### ***8.4.1 Competing Emotional and Cognitive Subsystems***

As presented in the Theories section, recent psychological theories on morality as well as neuropsychological models (e.g., the Somatic Marker Theory) claim that emotions are central for moral judgment and behavior. Following this "affective revolution" in psychology and cognitive sciences, many (early) neuroscientific studies were dedicated to the question of whether moral judgment and behavior is guided by reason or emotion.

One of the first studies investigating which brain regions are involved in moral judgment was the study by Greene and colleagues published in 2001. In this study, participants were presented with two types of dilemmas. One type is represented by the "trolley dilemma" and the other by the "footbridge dilemma". In the trolley dilemma, the participant is asked to consider the following situation: A runaway

trolley is quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman. If one does nothing the trolley will proceed to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single workman. After presenting this story, the participant in the experiment is asked to respond whether it is appropriate to hit the switch to avoid the deaths of the five workmen. In the footbridge dilemma the situation is slightly different. Again, a runaway trolley is heading down the tracks toward five workmen who will be killed if the trolley proceeds on its present course. The participant now has to imagine being on a footbridge over the tracks with a stranger and in between the approaching trolley and the five workmen. The only way to save the lives of the five workmen is to push the stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die as a result, but the five workmen will be saved. After presenting this situation, the participants again are asked to respond whether it is appropriate to push the stranger onto the tracks to save the five workmen.

By comparing neural activity during reasoning about these two types of dilemmas, Greene et al. (2001) found that reasoning about dilemmas that are emotionally engaging such as the footbridge dilemma (i.e., personal dilemmas or dilemmas in which physical harm is caused to another person directly by the agent) as compared to dilemmas that are less emotionally engaging such as the trolley dilemma (i.e., impersonal dilemmas or dilemmas in which physical harm is caused to another person only indirectly) activate the medial prefrontal cortex, the PCC, and the PSTS.

In a later study, Greene et al. (2004) further investigated how people solve particularly difficult personal moral dilemmas. An example for a very difficult personal dilemma is the “crying baby dilemma” that is used to bring cognitive and emotional processes into tension. In this dilemma, the participant has to decide whether it is appropriate to smother his or her own crying baby to save his or her life and the lives of other refugees hiding from enemy soldiers in a basement. Participants usually answer very slowly when presented with such a dilemma and do not reach a consensus on this issue.

The comparison of neural activity during utilitarian and non-utilitarian decisions (i.e., neural activity when participants decided that smothering the crying baby to save more lives is appropriate vs. neural activity when participants decided that smothering the crying baby is not appropriate) revealed increased activity in brain regions associated with abstract reasoning, conflict processing, and cognitive control such as the DLPFC and the anterior cingulate cortex (Greene et al. 2004). One interpretation of these results is that a conflict associated with such a difficult moral question is detected by the anterior cingulate cortex which then recruits control mechanisms and rational reasoning processes associated with neuronal activity in the DLPFC. These control processes help to resolve the conflict and to override prepotent emotional responses to make a utilitarian decision (see also Greene 2007; Greene et al. 2008).

To further investigate the role of emotion in moral judgment, Koenigs et al. (2007) tested a group of patients with VMPFC lesions. One of the most robust clinical findings in VMPFC patients is that they have blunt or flattened affects. For instance, in laboratory investigations VMPFC patients exhibit diminished autonomic arousal and subjective feelings in response to emotionally charged pictures and, according to their spouses, reduced feelings of empathy and guilt (Eslinger and Damasio 1985; Barrash et al. 2000). When confronted with moral dilemmas, these patients were more likely to choose a “rational” and utilitarian option (e.g., smothering a crying baby to save a group of refugees hiding from soldiers that normally would elicit a strong emotional response, see above) than healthy controls (Koenigs et al. 2007; Thomas et al. 2011; for similar results in a different sample of patients with VMPFC lesions, see Ciaramelli et al. 2007; but see also Kahane and Shackel 2008 for methodological problems in the study of utilitarian and non-utilitarian moral judgments).

However, it cannot be concluded that VMPFC patients in general decide more rationally. When engaged in real social situations involving frustration or provocation, the same participants exhibit exaggerated anger and emotional outbursts (Barrash et al. 2000). Using another experimental paradigm, namely the Ultimatum Game, Koenigs and Tranel (2007) found that patients with VMPFC lesions were influenced even more strongly by emotional reactions in their decisions. In the Ultimatum Game, two players are given a sum of money (\$100) and one opportunity to split it. The first player proposes how to divide the sum between each other, and the second player can either accept or reject this proposal. If the second player accepts, the money is split according to the offer. If the second player rejects, neither player receives anything. Therefore, a “rational” second player would accept any offer, no matter how low, because getting a low amount of money should be more rewarding and better than getting nothing at all. The “irrational” rejection of a low and unfair offer (e.g., when the offer is below \$20), in contrast, has been attributed to an impulsive reaction related to negative feelings such as anger and poor regulation thereof (Sanfey et al. 2003). Using this economic decision-making task, the authors demonstrated that patients with VMPFC lesions were more likely to make irrational choices and rejected low and unfair offers more often than healthy controls.

Together, the two different lesion studies show that VMPFC patients respond rationally in the face of abstract hypothetical scenarios related to the welfare of others, but irrationally in a real social setting involving their own self-interest. Although the precise role of VMPFC in moral judgment needs to be further investigated, it can be concluded that damage to this region disrupts the integration of emotion and reason in decision making. Further research is needed to investigate why a disruption of VMPFC function takes different forms in different circumstances (moral dilemmas vs. economic decision making) and whether emotion in general makes moral judgment better or worse (for further discussion and different explanations, such as a selective impairment of “prosocial sentiments” with a preserved capacity for anger or indignation, see reviews by Greene 2007; Young and Koenigs 2007; Moll and de Oliveira-Souza 2007; Young and Dungan 2011).

### ***8.4.2 Social Cognitive Processes and Mental State Reasoning During Moral Judgment***

As mentioned in the Theories section, the cognitive-developmental theory proposes that morality significantly relies on the way individuals understand and think about social situations. How people think about social situations is assumed to mature as a result of an active interaction of the individual with his or her social environment. Although this theory has been criticized and replaced by more recent models that place more emphasis on the role of emotion, neuroscientific work provides evidence that morality critically depends on a set of social cognitive abilities that allow people to take others' intentions, beliefs, and desires (or any kind of mental state) into account when making moral judgments.

Following the results of Rebecca Saxe and colleagues, the TPJ (mostly on the right hemisphere) appears to support important cognitive functions of mental state reasoning ("theory of mind") in moral judgment. These functions include the initial encoding of the agent's mental state (Young and Saxe 2008), the integration of that information (Young et al. 2007), spontaneous mental state inference (Young and Saxe 2009), and even post-hoc mental state reasoning to justify moral judgments (Kliemann et al. 2008; Young et al. 2011).

To investigate the impact of mental state reasoning in moral judgment, Young et al. (2007) used highly hypothetical scenarios in their studies. An example for such a hypothetical scenario is the following story: "Grace and her friend are taking a tour of a chemical plant. When Grace goes over to the coffee machine to pour some coffee, Grace' friend asks for some sugar in hers. The white powder by the coffee is not sugar but a toxic substance left behind by a scientist. Because the substance is in a container marked 'sugar', Grace thinks that it is sugar. Grace puts the substance in her friend's coffee. Her friend drinks the coffee and dies." (example from Young et al. 2007). By a systematic variation of outcomes (dying or not) and beliefs (she thinks it is sugar or she thinks it is toxic) the authors showed that neural activity in right TPJ was greatest for attempted harm; that is, in cases where protagonists were condemned for actions that they believed would cause harm, even though the harm did not occur.

Disrupting activity of the right TPJ by rTMS also disrupts the impact of mental state reasoning for moral judgment (Young et al. 2010a). Specifically, it reduces the role of intentions in moral judgment and increases, in contrast, the role of outcomes. For example, participants judged cases when actors intended but failed to do harm (including murder) as less morally blameworthy than cases in which harm was caused accidentally. However, TMS significantly reduced but did not completely eliminate the impact of mental state reasoning. In fact, there is evidence from many neuroimaging studies that the VMPFC, OFC, the temporal poles, and the PSTS also contribute to social cognition and mentalizing (Saxe et al. 2004; Amodio and Frith 2006; Mitchell 2009). Further evidence of a particular role of the VMPFC is given by a lesion study conducted by Young et al. (2010b) showing that patients with lesions in the VMPFC also judged cases with intended but failed harm as less blameworthy than controls.

In some research performed in our working group (Bahnemann et al. 2010), we investigated whether activity in the PSTS/TPJ region evoked by three tasks with increasing complexity (namely, detecting movements of bodies, making inferences concerning intentions, judging whether a behavior is morally good or bad) represents a common or distinct processes. We found an overlap of neural activity between all three tasks in right PSTS, but also a hierarchically increasing recruitment of the left PSTS and bilateral TPJ representing increasingly more complex processing of the social situation in the intention reading and moral judgment task.

In sum, these findings suggest that mental state reasoning represents a key cognitive component of moral judgment (i.e., moral judgments depend on information about agents' beliefs and intentions). The neural substrates that support this function, therefore, constitute an important part of the "moral brain network," in which the PSTS/TPJ and VMPFC are critical nodes.

### ***8.4.3 The Influence of Individual Differences in Moral Judgment Competence***

As already mentioned, to investigate the moral brain, some studies also investigated deviations and limitations in mental capacities thought to be relevant for moral judgment and behavior. These studies highlight the role of individual differences and provide direct evidence that particular abilities, such as empathy, perspective taking, and mental state inferences (in which patients differ from healthy controls) have a great impact on moral judgment and behavior.

The two studies by Koenigs and colleagues in patients with lesions of the VMPFC (Koenigs and Tranel 2007; Koenigs et al. 2007) discussed earlier, however, showed that a disruption of the affective/intuitive decision making component, on the one hand, improves utilitarian moral judgment, whereas on the other hand, economic decision making in the Ultimatum Game was impaired. Referring to the question of whether emotion makes moral cognition better or worse, Talmi and Frith (2007) stated that "The challenge, then, is for decision-makers to cultivate an intelligent use of their emotional responses by integrating them with a reflective reasoning process, sensitive to the context and goals of the moral dilemmas they face. If decision-makers meet this challenge, they may be better able to decide when to rely upon their emotions, and when to regulate them." (Talmi and Frith 2007: 866). Therefore, the question is not only which processes are involved in moral judgment but also how competently a decision maker can integrate the different (emotional and cognitive) processes sensitive to the context of the particular social situation he or she faces.

As mentioned in the Theories section, a current theoretical approach addressing this particular "intelligent use" of emotional and reasoning processes sensitive to the context of the specific social situation is the Dual Aspect Theory by Georg Lind. Referring to Kohlberg's notion of morality as an ability, Lind defines "moral judgment competence" as the ability to apply certain moral orientations in a consistent



and differentiated manner in varying social situations. Thus, social norms and values held to be virtuous in a culture or subculture are linked by means of moral judgment competence with everyday behavior and decision making (Lind 2008).

To investigate how individual differences in moral judgment competence are reflected in changes in brain activity during a moral judgment task, we conducted an fMRI study and measured neural activity while 23 participants made either moral or grammatical judgments. Participants were required to decide whether sentences were morally or grammatically correct or not. We correlated neural activity during these tasks with individual scores in moral judgment competence (Prehn et al. 2008). Individual moral judgment competence was measured using the Moral Judgment Test (MJT; Lind and Wakenhut 1980; Lind 2006, 2008; [www.uni-konstanz.de/ag-moral/mut/mjt-intro.htm](http://www.uni-konstanz.de/ag-moral/mut/mjt-intro.htm)).

The MJT confronts a participant with two complex moral dilemmas. In one dilemma (the doctor dilemma), for instance, a woman had cancer with no hope of being cured. She suffered terrible pain and begged the doctor to aid her in committing medically assisted suicide, and the doctor complied with her wish. After presentation of this short story, the participant has to indicate to which degree he or she agrees or disagrees with the solution chosen by the protagonist. After that, the participant is presented with six arguments supporting (pro-arguments) and six arguments rejecting (counter-arguments) the protagonist's solution which the participant has to rate with regard to its acceptability on a nine point rating scale ranging from -4 (highly unacceptable) to +4 (highly acceptable). Each pro- and counter- argument represents a certain moral orientation according to the six Kohlbergian stages.

In general, adult participants—in contrast to children or adolescents—prefer more elaborate arguments (i.e., adults rate more elaborate arguments as more acceptable than low level arguments) in line with having achieved a higher developmental stage of moral judgment. However, adult participants differ greatly in their ability to apply these orientations consistently especially when confronted with counter-arguments (i.e., arguments which are against their own opinion). This means that they rate more elaborate arguments as acceptable only when they represent their own opinion and reject all counter-arguments regardless of whether they are elaborate or not. The moral judgment competence score (C-score, the MJT's main score) reflects the ability to consistently or, in Lind's terms, competently apply a certain moral orientation and is calculated as an individual's total response variation.

By providing a measure of consistency, Lind's approach clearly goes beyond what we may ordinarily call "moral competence" as well as the Kohlbergian approach which focuses merely on moral orientations and the level of reasoning.

To our knowledge, the MJT is the only available test that provides a measure of moral judgment competence independent from a person's moral attitudes and values, in contrast to other instruments such as Kohlberg's Moral Judgment Interview (Colby et al. 1987), the Defining Issue Test (Rest 1974), the Sociomoral Reflection Measure (Gibbs et al. 1992), and the newly developed Moral Foundations Questionnaire (Graham et al. 2011), which all mostly assess individual moral orientations. The MJT has been proven to be a valid and reliable psychometric test.

For instance, moral judgment competence has been associated with responsible and democratic behavior. Translated in many languages, it also has been successfully used in scientific research (i.e., probing theoretical assumptions on moral development) and in evaluating educational programs (Lind 2008).

Contrasting neural activity during moral judgments with grammatical judgments, we found in line with the literature increased activation in the left VMPFC, the left OFC, the temporal poles, and the left PSTS for moral judgment. Regarding moral judgment competence, our sample of 23 participants showed a wide range of C-scores (maximum score=62.74, minimum score=5.55, mean=36.93, standard deviation=16.67). We correlated individual scores of moral judgment competence with neural activity during moral judgment and found that C-scores were negatively correlated with changes in BOLD activity in the right DLPFC during moral judgments contrasted with grammatical judgments. That is, participants with lower C-scores recruited the right DLPFC more than those with higher competence during moral judgment. Additionally, we investigated whether individual differences in moral judgment competence also modulate BOLD activity in the cerebral network engaged in moral judgment. An additional median split analysis revealed greater activity in the left VMPFC and the left PSTS in participants with comparably low moral judgment competence, specifically during identification of moral transgressions.

Finding a specific neural activation that reflects differences in moral judgment competence provides neuroscientific support for the Dual Aspect Theory by Lind. In the literature, greater neural activity in participants with lower ability in a certain cognitive task has been associated with compensation and an increased recruitment of mental resources (Rypma et al. 2006). As described earlier, moral judgment competence assessed with the MJT represents the ability to apply individual moral orientations in a consistent and differentiated manner in varying social situations. The increased activity in right DLPFC and left VMPFC/PSTS in participants with lower competence can thus be interpreted as reflecting higher processing demand due to a controlled application of moral orientations and an increased involvement of social cognitive and affective processes (such as mentalizing, estimating the value of possible outcomes of a behavior, and the experience of moral emotions) during the decision-making process (for extended discussion of the results regarding the brain regions involved, see Prehn et al. 2008; Prehn and Heekeren 2009).

Further neuroscientific evidence for a role of the right DLPFC in moral judgment and the implementation of morally appropriate behavior comes from a study using rTMS. Here also, a disruption of the right (but not the left) DLPFC reduces the subject's willingness to reject their partner's intentionally unfair monetary offers in the Ultimatum Game. Importantly, subjects were still able to judge the unfair offers as unfair. This indicates that the right DLPFC plays a key role especially in the implementation of fairness-related behaviors (Knoch et al. 2006).

Thus, both our own study and the rTMS study provide complementary evidence that there are specific brain regions crucial to the execution of morally appropriate behavior (see also Fecteau et al. 2007a, b; Knoch and Fehr 2007; Knoch et al. 2008).

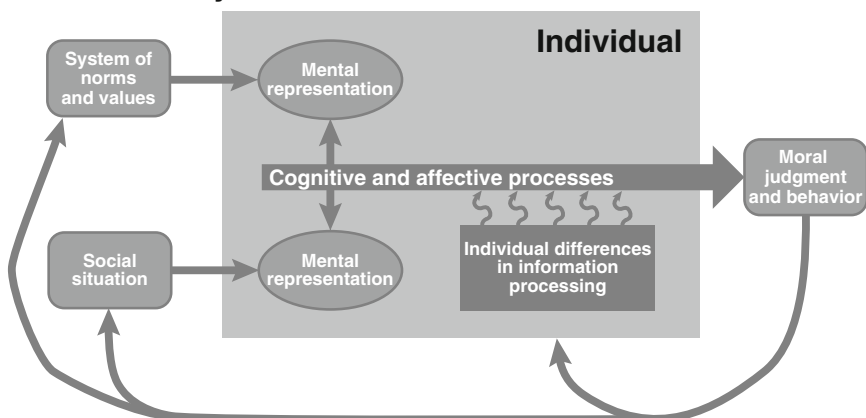
## 8.5 Conclusion

In this chapter, we took a look at current psychological models on moral judgment from a neuroscientific point of view, specifically introducing neuroscientific methods (clinical studies, fMRI, and neurostimulation) as powerful tools to investigate the processes underlying moral judgment and behavior in the human brain.

Since the availability of these tools, a multitude of studies have been conducted to investigate how and where morality is represented in the human brain. These studies used a variety of tasks and experimental paradigms, which are more or less ecologically valid, to investigate numerous different aspects, such as the role of emotion in moral judgment (see Sect. 8.4.1), social cognitive processes (Sect. 8.4.2), or the impact of inter-individual differences (Sect. 8.4.3).

What can we learn from these studies conducted so far? First of all, complex tasks, such as judging whether a certain behavior is “wrong” with regard to moral conventions, recruit a network of distributed brain regions supporting various subsystems and a number of different processes. For example, when confronted with a difficult moral dilemma, the situation presented to participants of a study needs to be represented in working memory. Simultaneously, accepted standards for social and moral behavior (i.e., norms and values considered to be virtuous in a culture or subculture) have to be retrieved from long-term memory. The presented behavior will then, subsequently, be evaluated with regard to these (re-)presentations. This evaluation process involves cognitive and affective sequences alike. Processes and mechanisms at the “cognitive” end of the spectrum, for example, might include social cognition, providing an understanding of the social situation and taking into account the cultural context and the intentions of the people involved. The individual would have to infer whether an actor did what he or she did on purpose or not. Processes at the “affective” end of the spectrum include the feeling of emotions such as guilt, sympathy, shame, anger, or disgust, when people are harmed and social norms are violated. The processes involved can be rational and accessible for conscious reflection (e.g., the controlled application of legal rules) or can be more subconscious, intuitive, and automatic. Depending on the circumstances (the emotional content of the situation, a person’s involvement, the necessity of reaching a utilitarian decision, etc.), either the cognitive or the affective aspect dominates in the decision-making process. For situations containing issues of life and death, which immediately affect the survival of an individual or his or her successful reproduction (e.g., in the case of consensual sex between siblings), judgment mechanisms might have evolved as a result of biological adaptation over the course of evolution and go on to proceed automatically and effortlessly without conscious reflection. Such mechanisms are suggested in the social intuitionist model by Jonathan Haidt. In contrast, coming to the conclusion that downloading music files illegally from the internet is harmful to society, for example, requires more abstract reasoning processes as proposed by the Kohlbergian model. As psychological theories and our own studies suggest, individual differences in information processing can also influence judgment processes and moral behavior. Individuals, for instance, considerably

## Culture / Society



**Fig. 8.1** A working model of moral judgment and behavior comprising different information processing modules described in this chapter. The judgment process includes cognitive and affective processes, which might be either more rational/conscious or intuitive/automated. In addition, information processing is modulated by individual differences such as the developmental stage of reasoning, moral judgment competence, or empathy. Finally, moral judgment and behavior have, in turn, an impact on the individual, his or her mental representations, as well as on future social interactions and the system of norms and values

differ in moral development, moral judgment competence, and empathy. Finally, moral judgments taken and the (moral or immoral) behavior shown could have, in turn, an impact on the individual (specifically on his or her experiences and mental representations), as well as on future social interactions and the society (e.g., when new laws are adopted to protect people which are treated immorally).

Taking all these different processes and aspects into consideration, we cannot expect to find morality located in a specific and distinct brain area (“a moral center”). The current model of the brain is an interconnecting networking model of information processing that integrates different kinds of information. That is, the “(moral) brain” can be broken up into several modules whose functions originally have nothing to do with morality (e.g., emotion, social cognition, cognitive control, etc.). By employing neuroscientific methods, the different “brain modules” or “process units” contributing to moral judgment and behavior can be (re-)presented, “visualized,” relatively analyzed, manipulated, selectively impaired, and (potentially) also modulated and trained. Neuroscience, thus, helps to imagine how processes are organized and grants a potential opportunity to alter human brain processing if something “is going wrong.” The goal of an empirically informed neuroscience of ethics is to integrate the various sets of subsystems contributing to moral judgment and behavior based on solid hypotheses stemming from all kinds of research fields. For a simple graphic idea of how the brain organizes moral processes, see Fig. 8.1.

A neuroscience of ethics, with its existing theoretical models, in our view, is highly beneficial to psychological studies and the field of psychology as such, both on a theoretical and practical level. We have taken the first steps toward “neuro-morality tests” and “neuropsychotherapy,” although we are still far from individual-based testing and neuropsychological assessment in forensic and pedagogical settings. One example of this endeavor is our study (Prehn et al. 2008) on the neural correlates of moral judgment competence (as defined following the Dual Aspect Theory by Lind). The data presented in this study strongly support the notion that morality can be considered as both a capacity and in terms of individual differences in the ability to apply frequently conflictual moral standards in a consistent and differentiated manner in varying social situations. It should be noted that it is presently unclear how exactly moral judgment competence, as measured by the MJT, maps on other cognitive abilities such as general intelligence. Future studies will have to address this question together with the question of how individual differences in moral judgment competence modulate processing during other kinds of tasks or day-to-day behavior. In addition, research is needed to investigate the possibility of changing one’s “ethics” through methodological training.

We hope that the review of neuroscientific studies on morality demonstrated that neuroscientific empirical research, to say the least, helps to disentangle the different processes involved in moral judgment and behavior. Thus, neuroscience helps to test and verify numerous assumptions made in psychological theories and can give an empirically informed opinion on long-time philosophical discussions aiming to differentiate between hotly debated concepts such as “intentions vs. consequences,” “intuition vs. reason,” or “morality of justice vs. care ethics.” However, and as we also emphasized in this chapter, one should not overlook the limitations or pre-conditions of neuroscientific methods and should always keep in mind the specific part of the broad philosophical notion of morality or ethics we are talking about at a given point in time.