Neural correlates of the motivation to be moral

Nunspeet, F. van

Chapter 1

General Introduction

This dissertation addresses a well-known but vast topic: Morality. Previous research has revealed that it is important for people to be moral. Nevertheless, they may sometimes commit immoral acts. In this dissertation, I take a social psychological perspective from which I examine when and why people become motivated to do what is right. I study whether people tend to adhere to their own moral values, and whether their moral behavior is affected by the presence of others. Moreover, by borrowing research methods from neuroscience, I aim to unravel some of the brain processes involved in this motivation to be moral.

Previous Research on Morality

Researchers across scientific disciplines, who examine different aspects of morality, work on the assumption that people have an innate sense of what is right and wrong. In fact, some of these researchers even argue that moral behavior is not unique for humans, but reflects a more basic concern for the well-being of others, that we share with some animals. For example, De Waal studied aspects of morality in chimpanzees, bonobos and capuchin monkeys. Results of his studies revealed that such animals show fairness concerns: When precious goods – such as attractive food items – are not equally distributed, they show signs of resentment (Brosnan & De Waal, 2003). Moreover, they comfort each other in distress and cooperate with other individuals in need of help, even if there is no immediate gain for the self (De Waal & Berger, 2000; see also De Waal, 1996). The fact that such indications of cooperation and empathy are found in primates (as well as other animals, such as elephants) is often interpreted as evidence that moral behavior represents a very basic and almost instinctive tendency – also for humans.

In the study of human behavior, developmental psychologists have theorized about how morality is established in childhood and develops through adolescence and adulthood (e.g., Kohlberg, 1969; Piaget, 1965). More recently, neuroscientific researchers have examined the effects of damage to (prefrontal) parts of the brain and have shown that such impairments are associated with immoral conduct and unethical decision making (e.g., Anderson, Bechara, Damasio, Tranel, & Damasio, 1999; for a review see also Moll, Zahn, De Oliveira-Souza, Krueger, & Grafman, 2005). These approaches thus also suggest that people have an intrinsic sense of morality (or a so-called moral intuition; see for example Haidt, 2001). Variations in

moral behavior seem to stem only from differences in the extent to which morality is developed in childhood or impaired due to physical restraints in the brain.

One could thus argue that people do not need explicit guidelines for what is the right thing to do, as they know this intuitively. This resonates with the consensus among researchers that moral principles are universal and fundamental to who we are. Yet, we are confronted with people's immoral acts on a daily basis: Every news website and –paper contains examples of people lying, stealing and cheating. Knowledge of the person who committed such an immoral act may surprise us. The people who are known for their good intentions, can still decide to act immorally. Likewise, research shows that the same individuals may show moral as well as immoral behaviors at different points in time (e.g., Monin & Miller, 2001). Why is this the case?

A Social Psychological Perspective on Morality

Prior attempts to answer this question have mainly investigated why people transgress moral norms. In line with the assumption that moral behavior is a natural tendency, such transgressions can be attributed to deficiencies in personal moral development or to cognitive limitations preventing people from showing 'regular' moral behavior. In this dissertation, I take a social psychological approach. I work on the notion that it is 'normal' for individuals to shift their moral behavior across situations or over time. I explicitly study these variations, focusing on *situational* features that induce moral behavior, as a starting point to increase our understanding of why and when people *adhere* to moral norms. Thus, the central aim of my research is to uncover which social mechanisms enable people to behave in line with their (and other people's) moral values, and how this affects the way they approach different situations. I argue that, by using this approach, we will gain a better understanding of how moral behavior can be stimulated by situational features. This can help bring out the best in people, regardless of their individual differences. To achieve this goal, I address three questions in the current dissertation: (1) Do people tend to act in ways that are considered moral? (2) How important is it for them to be perceived as moral by others? (3) How much do they care whether or not they succeed in behaving according to their moral values?

Examining these questions from a social psychological perspective means that I take into account the impact of how people see themselves, how they are judged by others, and to which social group they belong. In addition to examining social psychological factors in explaining displays of moral behavior, I use neuroscientific and psychophysiological indicators that may reveal the cognitive and affective mechanisms underlying such behavior. Combining these different approaches makes it possible to go beyond self-reported statements about what people say they will do. This also allows me to examine any discrepancies between the way people actually behave, and what they explicitly report. Going beyond prior work, my aim is to reveal whether and how people act upon their moral values by examining cognitive processes associated with moral behavior.

Diverging Perspectives on Morality

To examine what motivates people to be moral, we first must know what "being moral" actually means. In books of law or religion, morality is often defined by specifying what is *not* moral. The origin of current notions about human rights and general behavioral guidelines ('though shall not steal') can thus be traced throughout history and converges across national contexts, cultures, and religions. When moral standards are not made explicit, we may however still be guided by our moral intuition: An undefinable but certain intuitive state that indicates that something is right or wrong (Haidt, 2001).

The central goal of moral behavior thus seems quite obvious: Doing what is right. However, how this takes form in a concrete manner or in a specific situation is much more ambiguous. You may have noticed that what you consider the right thing to do may differ, depending on particular circumstances, or the presence of other people. For example, you know that helping others is generally considered moral. Nevertheless, you may be more motivated to help your friends or family members than some stranger in the street. In a similar vein, you are likely to care whether others perceive you as a moral person. At the same time, opinions of others you consider relevant to yourself – such as your friends or family – are likely to matter more than opinions of people you do not know. As a consequence, whether or not you act upon general moral guidelines is likely to differ, depending

on who is affected by your behavior, or who are present to observe and evaluate your behavior.

Individuals deliberate about what would be the right thing to do, but so do groups, institutions and countries. To give an example, let us consider the Olympic Winter Games of 2014. Ever since the Olympic Committee announced that these Games would be hosted by Sochi –Russia, one could hear objections around the world. Several countries objected to the organization of an event that promotes peace and international cooperation through sports, by a country that is associated with limited civil rights such as rights of freedom of assembly and freedom of speech. Also in the Netherlands, there were fierce discussions about the decision of the government to send a large political delegation (in addition to members of the royal family) to attend the Games. According to protesters, this signaled the wrong message: Given the high moral standards concerning civil rights in the Netherlands, this country should not support an event organized and propagated by another country that violates such rights. Such debates thus raise questions about national moral values, how to (re)present those values, and how these will be perceived by other communities and countries.

The above example illustrates that there can be differences between groups in moral values on an international level. Debates about what is right may however also divide different groups within the same country. For example, in the Netherlands, Belgium, and France, discussions concerning the integration of Muslims invoke moral concerns. Norms posed by the Islam seem to oppose common societal practices in Western countries. This is the case for instance with the clothing habits that many Muslims endorse, such as wearing a burka or a hijab – a headscarf– for Muslim women. Wearing a burka in public was prohibited in France in 2011. A similar judicial proposition was discussed in the Netherlands as well. Wearing such clothing may be perceived in Western societies as degrading for women and as morally wrong because it could strengthen the segregation of Muslim and non-Muslim individuals. In contrast, Muslims see this as a sign of modesty and high moral standards. This illustrates that the same behavior (such as wearing a headscarf) can be considered the moral thing to do by some, while being

seen as immoral by others. In other words, it is not always easy to specify the 'right' thing to do because each group may have its own moral norms.

In a context where members of multiple groups are present, people can therefore question what would be the moral thing to do. When the former queen of the Netherlands went to Oman for a state visit, she wore a headscarf whenever she visited a mosque. She argued that she did this out of respect and regard for the country, its people and their religion. Several members of the Dutch government supported her judgment. However, there were also politicians who openly condemned her opinion and related behavior. This example thus also illustrates that debates about what is moral touch upon who we are as individuals, and how we see ourselves in relation to our groups (e.g., a political party, 'the Dutch'). They also concern our moral principles and values; how we want to portray ourselves to others; and how we want to be perceived by them. These are questions that are central to the current dissertation.

Morality and Group Inclusion

The importance of the people around us for how we think about ourselves and decide upon how to behave can be explained from a social identity approach. Social identity theory posits that people often perceive themselves and others as part of a group. Groups help people to define who they are, where they belong, and how they should behave (Tajfel & Turner, 1979). Being part of a group with whom one can share his or her social identity (e.g., "the Dutch", "social psychologists") is a way to validate one's self-views, and to establish and maintain one's self-esteem (see also Ellemers & Jetten, 2013). Groups thus can help people to establish a distinct identity: Groups each have their own norms which make them different from other groups. The norms and values within a group thus provide clear guidelines as to how individuals should behave in order to secure inclusion in that group. As a result, people tend to look for inclusion in a group with whom they can share their moral values and principles. Alternatively, they adapt their own values to the groups that are important to them. It thus depends on whether people want to belong to and identify themselves with a particular group (whether they consider this their 'ingroup') whether they adjust their behavior according to the groups' moral norms. This refines the idea that people's

behaviors are affected by 'social pressure' in general. People care primarily about adherence to norms within their ingroup, while it is less important for them to behave according to outgroup standards. For example, Dutch Muslim women who identify more strongly with their religious group than their nationality are more likely to adhere to the norms of their religious group (e.g., by wearing a headscarf) than the norms of their Dutch nationality (e.g., not wearing a headscarf).

Group norms and standards are particularly important when these relate to morality. As a member of a group, people are more inclined to adhere to ingroup norms when these are presented as "the moral thing to do", rather than prescribing what would be "the competent thing to do" (Ellemers, Pagliaro, Barreto, & Leach, 2008). People do this because they think they will receive respect from their fellow group members when they adhere to moral group norms (Pagliaro, Ellemers, & Barreto, 2011). Moreover, people identify more strongly with a moral than a competent group and are more proud to be member of groups that can contribute to their morality than groups that stand out for their competence (Leach, Ellemers, & Barreto, 2007).

Morality also seems to be the most important determinant of the impression we form of other individuals and groups. When encountering someone we do not know, we primarily search for characteristics indicating their morality (e.g., honesty, trustworthiness) rather than showing an interest in competence (e.g., particular skills, intelligence) or sociability (e.g., kindness, friendliness; Brambilla, Sacchi, Rusconi, Cherubini, & Yzerbyt, 2012).

Thus, both at an individual and a group level, people look for characteristics concerning morality –rather than information concerning other people's competence or sociability (Brambilla, Rusconi, Sacchi, & Cherubini, 2011). In fact, research shows that we are able to determine whether another person is trustworthy in less than a second. This happens even faster than making judgments about whether that person is attractive, competent, or nice (Willis & Todorov, 2006). In the process of gathering information about how moral someone is, special importance is attached to any negative behaviors. That is, we more likely to conclude that someone is immoral when s/he has done something wrong, than we are to conclude that this person is moral because s/he is always honest and reliable.

In other words, even for a person who is known for his or her moral integrity, a single act of immoral conduct can spoil this positive image, because immoral acts are perceived as more informative of someone's true character than moral acts (Skowronski & Carlston, 1987).

This is not only important when learning about someone else's moral characteristics and values, but also plays a role in the concerns people have about *themselves* being seen as moral by others. That is, if one's morality is called into question, then one's identity and sense of self is negatively affected. For example, when there is disagreement about moral values (as compared to material interests), or when a person is evaluated on his or her prior immoral behavior, people report increased negative affect and display a physiological threat response (Kouzakova, Harinck, Ellemers, & Scheepers, 2014; Van der Lee, 2013).

When others question their moral intentions or behaviors, people worry that they may lose respect or even will be excluded from the group. However, since the meaning of morality differs between groups and situations, it can be impossible to do what is right according to everyone. People may therefore focus primarily on doing what is right according to their own ingroup. Such ingroup norms may however also concern how one should behave towards members of another group (e.g., treating people from other cultures with respect). The intention to adhere to such ingroup norms may be relatively easy as long as interactions with outgroup members are hypothetical. But what happens when people are faced with an actual interaction with a member of another group? For example, when non-Muslim have to collaborate with a Muslim at work?

Morality in Intergroup Relations

As I explained above, morality plays an essential role in regulating individual behavior within a person's own group. It is however just as important in intergroup interactions. Accordingly, morality is often examined in such contexts. For example, Reed and Acquino (2003) revealed how intergroup conflict can be diminished by extending ingroup favoritism towards individuals representing different religions and ethnicities. That is, people show increased explicit moral regard towards outgroup members when they attribute greater importance to their moral identity (Acquino & Reed, 2003). Likewise, previous research has shown that

people's willingness to strive towards social equality between groups is enhanced when other ingroup members say this is an important moral ideal (rather than when they say it is a moral obligation; Does, Derks, & Ellemers, 2011). Evaluating or presenting people's identity or behavior in terms of moral values can thus enhance their moral intentions and acts. In other words, telling people that they should act according to what they think is the right thing to do may thus be used as an instrument to enhance moral behavior towards and between people. I assess the implications of the effects of such an emphasis on a person's morality, in the current dissertation. Specifically, I examine what happens when the implications of behavior of native Dutch (non-Muslim) individuals towards Muslim women are presented as an indication of their egalitarian values. I propose that reminding people that their behavior conveys their morality will stimulate equal treatment and motivate people to avoid displaying bias towards the Muslim outgroup.

Measuring Moral Behavior

Thus far I have discussed why it may be important for people to adhere to their own moral values and the moral norms within their groups. If people indeed want to be perceived as moral, this could cause them to emphasize the importance they attach to moral behavior because they think this may reflect positively upon the image others have of them. This may however not necessarily reflect their actual behavioral preferences. Nor does it predict how they would act in a specific situation, for instance when they do not realize that others are paying attention to their moral tendencies. In other words, people may deliberately respond in a socially desirable fashion when they think their moral image is at stake. This is a relatively common concern when interpreting responses to self-report questionnaires. Emphasizing the implications of people's behavior in terms of how moral they are could thus introduce measurement problems. Relying on self-reported intentions to assess people's responses may not capture their 'true' intentions, or their intentions may not correspond to their actual behavior. In this dissertation, I therefore used another type of measure to assess the motivation to be moral. I adapted an Implicit Association Test (IAT) to examine participants' behavioral responses that might reveal bias favoring their own ingroup (non-Muslims) over members of a relevant outgroup (Muslims).

The IAT was first developed by Greenwald, McGhee, and Schwartz (1998) to assess the strength of (automatic) associations between target concepts and different attributes. This test assumes that people find it easier to quickly connect concepts that they implicitly relate to each other. You can imagine how this works when you are asked to couple a concept, such as "flowers", with words like "fun" or "kind" (i.e., attributes). Making such connections should be relatively easy because "flowers", as well as words like "fun", both have a positive connotation in our mind We thus associate one with the other, because they are what is called *congruent*. Likewise, it should be relatively simple to couple a concept such as "bugs" with words like "pain" or "fear". In this case, the association is easily made because both the concept and the words have a negative connotation. However, things are likely to become more difficult when you try to couple "flowers" with "pain", or "bugs" with "fun". This is because these concepts and words do not have the same connotation - a positive word has to be coupled with a negative concept - and are thus *incongruent*. They are therefore not easily associated with one another.

An IAT is based on these associative mechanisms. It is a reaction time task during which participants are asked to press one key as quickly as possible when they see a particular word or picture. In one part of the task, they are asked to respond with the same key to both pictures or names of "flowers" and positive words (e.g., "fun"). They are asked to press another key for both pictures and names of "bugs" and negative words (e.g., "pain"). This procedure is used to assess participants' performance on congruent trials. In another part of the task, the pairing becomes less intuitive. Here, participants are asked to respond with the same key to both "flowers" and negative words. Another key has to be used to indicate both "bugs" and positive words. These instructions are used to assess participants' performance on incongruent trials. To the extent that people are more inclined to associate flowers with positivity and bugs with negativity, they should respond more quickly to congruent trials than incongruent trials. Thus, the difference in their reaction times on incongruent compared to congruent trials reveals the strength of their implicit associations. This is what is called the IAT effect. It indicates the extent to which people find it more difficult to associate one concept (e.g., "flowers) with negative rather than positive words, and another

concept (e.g., "bugs") with positive rather than negative words. The difficulty of making such associations is revealed in increased response times. In this way, the IAT effect can reveal people's negative bias towards all kinds of manner of target concepts, including bugs.

The example of flowers and bugs illustrates the principles on which the IAT effect is based. However, the test has most often been used to assess implicit negative bias towards different groups of people in society, in studies concerning prejudice. In this case, you are also asked to couple a concept with positive words. But this time, the concept is not "flowers", but represents a social group, for instance native Dutch people. As indicated above, people are concerned with having a positive social identity. Hence, they are likely to think more positively of groups associated with the self (ingroups) than of other groups (outgroups). The groups to which they belong (and Dutch participants can be seen to belong to the group "the Dutch") is likely to have a positive connotation. In comparison, people are more likely to have negative connotations with an outgroup, such as immigrants. When performing the IAT, responding with one key to the concept "native Dutch" and positive words may thus be relatively easy, as is responding with another key to the concept "immigrants" as well as negative words, as these represent congruent associations. In contrast, responding with a single key to the concept "native Dutch" and negative words is likely to be more difficult –and will therefore take more time– just as responding with another key to the concept "immigrants" as well as positive words (incongruent associations). The IAT effect (i.e., the difference in response times between incongruent and congruent trials) in this case reveals the extent to which it is more difficult to associate one's ingroup (e.g., "native Dutch") with negative rather than positive words, and an outgroup (e.g., "immigrants") with positive rather than negative words. In other words, the IAT score can reveal people's implicit negative bias (prejudice) towards immigrants.

The target concepts "native Dutch" and "immigrants" are an example of concepts that can be used in an IAT. In the United States, the IAT is often used within a racial context. Such a 'race IAT' consists of stimuli (such as photographs) representing people with a white or dark toned skin color. Explanations for white people's tendency to reveal a negative bias towards black people are diverse. People

with a dark skin tone may be seen as outgroup members by people with a white skin tone. The differentiation between these two groups may thus reveal positive associations with the ingroup and negative associations with the outgroup. However, in the case of the 'race IAT' other explanations could also be offered for this pattern of associations. For example, stereotypes of people with a dark skin tone may more often be negative rather than positive. Think for example about stereotypes concerning criminal records and aggression. As a result, the physical features of a black man's face may be perceived as more threatening than the physical features of a white man's face, which could cause negative rather than positive associations with this type of stimulus. All these explanations could thus explain the emergence of negative bias on the IAT, against people with a dark skin tone.

In the current dissertation I use different target concepts in the IAT because of two reasons. First, negative stereotypes concerning people with a dark skin color as well as discrimination against this group are less common in the Netherlands. Such a 'race IAT' is thus less relevant to assess in a Dutch research population. Second, I attempt to rule out some of the additional explanations for a negative outgroup bias –besides the explanation of one's social and distinct social identity. The IAT target concepts I use in this dissertation are "women without a headscarf" and "women with a headscarf". Women without a headscarf represent native Dutch, non-Muslim women. These women are similar to my research participants and thus are likely to be seen as ingroup members. Women with a headscarf represent Muslim individuals. These women are different from my research participants and thus are likely to be perceived as outgroup members. As I indicated in the first part of this introduction, the integration of Muslims is a current topic of debate in the Netherlands. This debate to an important extent addresses clothing habits, such as wearing a headscarf, in public places or functions. Measuring people's negative bias against Muslim women (i.e., women who wear a headscarf) is thus more relevant for research in the Netherlands. Furthermore, I pretested the photographs of the faces of these women (i.e., the stimuli in the IAT) on different characteristics. Examples are perceived kindness, honesty, intelligence, and attractiveness. Results of this pretest showed that both the women with and

without a headscarf are perceived as equal concerning these characteristics (see Appendix A of this dissertation for more details). A negative bias against women with a headscarf –as revealed by this IAT– can thus not be explained by any negative associations (related to such characteristics) with the stimulus materials as such. Importantly, the Muslim women presented in the IAT were only perceived as *different* from the research participants –but not in any way more negatively or less positively than the non-Muslim women. If I find a difference between positive and negative associations with Muslim and non-Muslim women this can therefore only be attributed to the fact that the women with a headscarf are being perceived as different from the research participants – i.e. as outgroup members. In other words, the use of these stimulus materials implies that any negative bias against Muslim women that is revealed by the IAT, can only be attributed to the fact that these individuals are seen as representing another (out)group.

Emphasizing the Implications of One's Behavior

In my research, I use the IAT as an indicator of people's negative bias, or prejudice, towards outgroups. Some would propose that the associations between groups and positive and negative attributes that people make during the IAT, are made easily and quickly because they occur automatically. However, prior research has revealed that IAT performance is malleable: The fast elicited response to associate some concepts and attributes are not automatic, but can be adapted. That is, participants can deliberately influence their performance by using strategies that diminish the difference between response patterns on congruent and incongruent trials. This is the case, for instance, when they are informed about how their bias will be measured (Fiedler & Bluemke, 2005). Likewise, IAT responses are adapted when people are explicitly motivated to enhance their self-image or to emphasize their positive relationship with other individuals (for an overview see Blair, 2002). Thus, using an IAT, it is possible to examine whether people adjust their performance when motivated to do so. In the current dissertation I examined whether participants performed differently when they were reminded of the moral implications of their behavior during this task. That is, I examined whether people showed less implicit bias against Muslim women when they thought that the test would reveal whether they treated Muslims and non-Muslims equally (e.g., their

moral values concerning egalitarianism and discrimination), rather than merely being good at quickly processing information and learning to make new associations.

Because of the stimuli used in an IAT, performance on the test can relatively easy be seen as indicating prejudice, and thus perceived as a measure of moral values. However, at the same time, the test is a reaction time task in which participants are asked to sort different types of stimuli according to changing rules. The faster and more accurately participants respond, the better their performance. Thus, it would be equally plausible to see the IAT as a test of people's ability to perform well on this task. In other words, the IAT can be perceived both as a measure to detect social bias against an outgroup, *and* as a measure of one's competence. My aim is to examine whether people respond differently during the IAT depending on which of these task implications is emphasized. This allows me to investigate whether (and how) people adjust their behavior when they think their performance can indicate their moral values concerning egalitarianism.

In most of the studies reported in this dissertation (i.e., Chapters 2 through 5), the IAT is used to assess behavioral responses. This approach extends previous research concerning the importance of morality for people's self-views and social identity, which has mainly relied on explicit self-report measures. Since people may adjust their deliberate responses on a self-report questionnaire to convey what they think is perceived as moral by others, their answers may not necessarily reflect the way they will actually respond in situations where moral concerns play a role. Assessing people's moral responses in a less explicit way, by using this IAT, can thus provide insight in whether people actually behave in line with relevant moral values. In addition to assessing task behavior to reveal implicit bias, I use psychophysiological and neuroscientific research methods to increase our understanding of the cognitive processes that underlie people's adherence to their moral values.

The Added Value of Cognitive Neuroscience

An important additional aim of the research reported in this dissertation is to examine the cognitive and neural mechanisms that underlie people's motivation to act according to what is moral and to be perceived as moral by others. Previously, I

explained how behavioral performance on an IAT can give us more information about a person's 'true' behaviors. In addition, I aim to uncover *how* people monitor and adapt their behavior to achieve adherence to moral values. Understanding the cognitive processes that help people to behave in a moral manner may expand our knowledge of the mechanisms needed to behave morally. It may reveal whether people *initiate* their behavior in a different way when they think this is indicative of their moral values (as compared to, for example, their competence). People may for instance pay more attention to other people's skin color when they have just been informed about discrimination rates. As a result, they may more quickly detect someone with a different ethnic background which will help them to act in an unprejudiced manner. On the other hand, the cognitive mechanisms may reveal increased vigilance to errors, which may help people to *adjust or redirect* their responses to avoid displaying signs of possible immoral behaviors when they want to appear moral.

This type of behavioral initiation or correction is likely to occur outside of one's conscious awareness. In a job interview for instance, an employer may be focused on the applicant's gender because he read a report the day before about the under-representation of women in business organizations. At the same time, the employer may not be aware of his increased attention to that aspect of the applicant. He may not even consciously remember reading that report. When asked to verbalize his considerations, the employer may thus be unable to report that he was more focused to the applicants' gender. In fact, even when the employer was aware that he was more attentive to the gender of the applicants that day, he may not want to disclose this for fear of revealing gender bias. In other words, people may not be *able* to tell us about the cognitive processes that they recruited in order to behave in a particular way. And even if they are able, they may not be *willing* to tell us about those processes.

Neuroscientific research methods can help solve such problems. Methods used in cognitive neuroscience have proven to be effective in gaining insight in processes such as enhanced or decreased attention. Using such measures can thus reveal additional information about the mechanisms underlying people's actual behavior. They make it possible to capture automatic and/or unconscious response

tendencies elicited by moral situations. Additionally, such neuroscientific measures provide an unbiased perspective on what actually happens during task performance, as these indicators are not sensitive to people's supposedly heightened social desirability to comply to moral expectations.

Extending the Cognitive Neuroscience of Moral Reasoning

Cognitive neuroscientists have already begun to shed light upon the cognitive and neural mechanisms associated with moral reasoning and decision making. For instance, previous research has examined how people reason when they are confronted with a moral dilemma and asked to decide how they would behave in such a scenario. A famous example concerns the so-called 'trolley dilemma' in which people are asked whether they would sacrifice one person's life in order to save five other individuals (Foot, 1978; Thomson, 1985). Neuroscientific research has revealed that brain networks associated with both cognitive as well as emotional processes are involved in such moral reasoning (e.g., Greene, Nystrom, Engell, Darley, Cohen, 2004). Moreover, research concerning the judgment of moral and immoral acts has revealed that people are highly sensitive to the detection of moral transgressions which may be related to the instant emergence of moral emotions such as disgust (e.g., Borg, Lieberman, & Kiehl, 2008; Schnall, Benton, & Harvey, 2008). Such studies have thus focused on the mechanisms underlying people's individual ability to reason about and decide what is and what is not moral. However, as I have explained above, social contexts may affect what can be considered the moral thing to do. Likewise, different situations may affect whether people actually behave according to what is perceived as moral. These social factors are often neglected in cognitive neuroscience, as much of the research in this tradition focuses on establishing universal response patterns. Nevertheless, I argue that moral behavior is likely to shift across different contexts, depending on the social concerns that are raised. Additionally, knowing right from wrong and being able to make moral judgments may differ significantly from people's actual moral intentions, motivations, and subsequent behavior. Thus, to gain better understanding of people's motivation to adhere to their own moral values, and how they enact this motivation, I will investigate the mechanisms underlying actual

moral behavior (i.e., IAT performance), and how these are affected by different social contexts.

Multiple Research Methods

Besides using self-reports and measuring behavioral responses on the aforementioned IAT, I used three different research methods in the studies reported in this dissertation: Skin conductance, EEG, and fMRI.

Skin conductance.

Skin conductance indicates electrodermal activity representing activation in the sweat glands, measured at the skin surface of our hands. Skin conductance relates to so-called "psychologically induced sweating" (Dawson, Schell, & Filion, 2000, p. 202). People automatically sweat when they experience emotions, when they become aroused, or when their attention is increased. Measuring the tonic level of skin conductance can thus be used as an unobtrusive way to examine general states of arousal and alertness. Moreover, phasic skin conductance responses (SCRs) can be elicited by different characteristics of an occurring event. In psychological experiments this may be a particular stimulus that is new, intense, or has an emotional impact. Skin conductance is an automatic response generated by the sympathetic nervous system, a process that thus cannot easily be adapted by the participant for self-presentational reasons. Additionally, variations in skin conductance can be measured *while* participants receive relevant information. I am interested in whether people care about succeeding in behaving according to moral norms. Skin conductance is thus a valuable measure to detect how people (physically) respond to information indicating that they are, or are not, as moral as others.

EEG.

EEG is the abbreviation of *electroencephalogram*, which is an indicator of brain activation measured across the scalp (e.g., Luck, 2005). EEG has a relatively low spatial resolution: It is usually unclear from which brain region the activity originates, because it is measured at the scalp. This noninvasive neuroimaging technique does however have a high *temporal* resolution: Evoked responses in brain activation can be measured within milliseconds after a stimulus is presented on the screen or when a response is given.

An EEG can be used to monitor ongoing brain activation during a complete experiment. From this EEG, we can extract responses evoked by particular events –so-called event-related potentials (ERPs; Luck, 2005). Using ERPs, it is thus possible to gain insight in the (ongoing) cognitive processes associated with particular parts of the experiment. For example, ERPs around a given response can inform us about the preparation of and the reaction to that response on a cognitive level. This thus complements the actual behavioral response that can only indicate for instance what people decide to do, or how long it takes for a participant to make this decision.

Besides examining ERP's during task responses, we can also investigate how different stimuli are processed in the brain. In the IAT I developed for the research in this dissertation, the target concepts are presented by photographs. Specifically, the target concept "ingroup" is represented by photographs of women without a headscarf. The target concept "outgroup" is represented by photographs of women with a headscarf. Using ERPs, it is possible to detect that these two types of photographs are differentially processed in the brain. In addition, I can examine whether the ERP modulations associated with viewing ingroup and outgroup members are affected by people's motivation to perform in line with their moral values. In addition to a computation of the behavioral responses on the IAT (i.e., response latencies and the amount of accurate responses), ERPs can thus reveal the attentional processes associated with such task performance.

fMRI.

In contrast to EEG, functional Magnetic Resonance Imaging (fMRI) is a neuroimaging technique that has a relatively low temporal resolution but a high *spatial* resolution – it reveals which brain areas are activated (e.g., Huettel, Song, & McCarthy, 2004). Although also noninvasive, MRI is used to visualize internal physical tissue. Moreover, within the brain, *functional* MRI is used as an indicator of brain activation during task performance. Using the blood-oxygen-level-dependent (BOLD) response, differences in deoxygenated blood levels are measured. Performing a task elicits specific cognitive demands, such as increased attention. For such cognitive demands an increase in energy is needed in particular parts of the brain. One of those sources of energy is oxygen. Release of this oxygen

increases the level of deoxygenated blood which can be detected using magnetic resonance. This is thus used as the indicator of brain activation (Huettel et al., 2004). The whole brain can be visualized using MRI and the BOLD-response can be measured to localize activation in specific brain regions. It is thus possible to compare the degree and location of brain activation associated with different parts of an experimental task. This also implies that we can detect activation in *sub*cortical regions of the brain (that are located deep in the brain), including structures associated with primary affective responses. In addition to behavioral and self-report measures, but also extending information gathered with ERPs, fMRI can thus inform us about the brain regions involved in moral task performance.

Overview of the Dissertation

With the research reported in this dissertation, I address three different research questions. In Part I, I examine whether people tend to act in ways that are considered moral. in Part II, I investigate how important it is for people to be perceived as moral by others. Finally, in Part III, I focus on how much people care whether or not they succeed in behaving according to their moral values.

In Part I, I examine whether people tend to be more motivated to show that they are moral than that they are competent. To be able to make this comparison, I present an IAT as either indicative of one's moral values concerning egalitarianism and discrimination, or of one's ability to learn new tasks and to quickly process information. In Chapter 2, I examine whether participants' task performance is affected by this difference in emphasis on specific task implications. Specifically, I test the prediction that when the moral implications of the task are stressed participants show a weaker negative bias against Muslims than when the competence implications are emphasized. Additionally (using EEG in Chapter 2, and fMRI in Chapter 3), I examine whether the moral test implications enhance participants' attention towards different aspects of the task. In other words, these studies aim to reveal whether stressing the implications of one's behavior in terms of one's moral values causes people to adjust and direct the focus of their attention during task performance.

In Part II, I examine how important it is for people that others think they are moral. In this part of the dissertation, the implications of the IAT are again presented in terms of one's moral values or competence. Additionally, participants are led to believe that their performance on the IAT is being monitored and evaluated by someone else. In Chapter 4, I examine whether people show their motivation to be a moral group member by inhibiting their bias against Muslims when an ingroup rather than an outgroup member is evaluating their performance. In this chapter, the evaluator is a non-Muslim individual (who's gender is matched with that of the participant). In one condition, she is presented as someone with the same group membership as the participant. This is achieved with very minimal instructions (also referred to as the 'minimal group paradigm'; Tajfel, 1970). The participant is told that their evaluator has the same personality type as they do and that s/he is thus an ingroup member. In another condition, the evaluator is presented as someone representing the other minimal group. Participants in this condition are thus told that their evaluator has another personality type than they do and that s/he thus can be considered an outgroup member. As in Part I, I thus examine whether people adjust their behavior and increase their attention towards the task in case the moral implications are emphasized. In addition, I test whether participants are more inclined to do this when an ingroup member, rather than an outgroup member, is evaluating their behavior.

In Chapter 4, participants' IAT performance is thus monitored by a non-Muslim individual who is introduced as someone with the same or another personality type as the participant. Thus, the evaluator is introduced as someone who is similar to or different from the participant based on an implied personal feature. Nevertheless, the person evaluating participants is always the same man or woman, and the evaluator's visible appearance is always the same as that of the participant. In Chapter 5, I examine whether people's moral behavior towards an outgroup (i.e., appearing unprejudiced) is affected when they are being monitored by someone who can be seen to represent the target outgroup in the IAT: A woman with a headscarf. Such an evaluator can thus be seen as an outgroup member. In principle, being seen as moral by outgroup members should be less important than being seen as moral by ingroup members. This might imply that

participants should not be motivated to appear moral towards their Muslim evaluator – because she is an outgroup member. On the other hand, since the moral behavior assessed in this research is people's bias against Muslims, a Muslim evaluator (representing the target group against whom bias might be revealed) could still have an impact on participants IAT performance – albeit for different reasons (see also Lowery, Hardin, & Sinclair, 2001). I thus examine the effects of being evaluated by a Muslim woman on moral task performance. Additionally, I compare the impact of being monitored by this Muslim evaluator, depending on whether she is presented as an ingroup or an outgroup member based on the previously described minimal group membership. Specifically, participants are informed that their evaluator either has the same or another personality type as they do. I thus also examine whether presenting an outgroup member (the target group representative) as a partial ingroup on another dimension (same personality type) helps people to reduce prejudice against the target outgroup.

In Part III, I again address people's motivation to act according to what is considered moral. But here I go one step further. I focus on how much people care whether or not they *succeed* in behaving according to their moral values. In Part III, I aim to extend the findings of Part I, in which I examine whether and how people's motivation to be moral affects their performance on a measure of bias against Muslims. In Chapter 6, participants are provided with feedback about their performance on this test. They either are confronted with information indicating that they are less moral (or less competent) than other participants, or with information stating that they are more moral (or more competent) than other participants. I examine the emotional and psychological impact of this information. Specifically, I measure self-reported affect and skin conductance responses as an indicator of physiological arousal. If people care more about being moral than competent, receiving negative information about their own moral behavior should be more distressing than being confronted with negative information about one's competence. Nevertheless, or even because of this reason, people may be more attentive to positive information about their morality since this may confirm their moral self-concept. In an additional fMRI study I further examine this prediction

and compare whether information related to one's morality rather than one's competence is processed as more self-relevant in the brain.

In the final part of the dissertation (Chapter 7), I integrate the findings presented in the three empirical parts. I discuss their implications and how this research contributes to current insights in social psychology and social neuroscience, and I consider the societal implications of my findings. Note that Chapters 2, 3, 4, 5, and 6 are prepared as separate journal articles. This results in some overlap in the theoretical background and method sections, but also implies that these chapters can be read independently.