



Universiteit  
Leiden  
The Netherlands

## Neural correlates of the motivation to be moral

Nunspeet, F. van

### Citation

Nunspeet, F. van. (2014, May 27). *Neural correlates of the motivation to be moral*. Kurt Lewin Institute Dissertation Series. Ridderprint B.V., Ridderkerk. Retrieved from <https://hdl.handle.net/1887/25829>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/25829>

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/25829> holds various files of this Leiden University dissertation.

**Author:** Nunspeet, Félice van

**Title:** Neural correlates of the motivation to be moral

**Issue Date:** 2014-05-27

Chapter 7

# General Discussion



In this dissertation, I examined three different research questions. In Part I, I tested whether people tend to act in ways that are considered moral. In Part II, I addressed the question how important is it for people to be perceived as moral by others. In Part III, I examined how much people care about whether or not they succeed in behaving according to their moral values. Additionally, I aimed to unravel the cognitive processes associated with these motivations. In this final chapter, I will discuss the conclusions that can be drawn from the research reported in the previous five chapters. First, I will review the behavioral findings observed in the three different parts of the dissertation. Then I will elaborate on evidence revealing the underlying processes associated with these behavioral results.

### **Behavioral Findings**

#### **Part I: Moral Concerns Cause Inhibition of Intergroup Bias**

Previous research has revealed that people explicitly report that they think it is more important to be perceived as moral than as competent (Ellemers, Pagliaro, Barreto, & Leach, 2008). One aim of the current dissertation was to examine whether people not only explicitly report this motivation, but actually tend to behave more according to their moral than competence values. To assess this, I presented native Dutch, non-Muslim, research participants with an implicit association test (IAT). This test is a measure of one's (implicit) prejudice towards a particular outgroup –in my research these were Muslim women. I framed this test as being able to show how moral or how competent people are. Specifically, participants were either informed that “this task can give an indication about your moral values concerning egalitarianism and discrimination”, or that “this task can give an indication about your ability to learn new tasks and to quickly process new information”. I thus examined whether the implicit bias people showed in their task behavior would be reduced to a greater extent when they were motivated to be moral than when they were motivated to show their competence. Results in Chapters 2, 4 and 5 (Studies, 2.1, 2.2, 4.1, and 5.1) revealed that people indeed are more motivated by their moral than their competence values. In case of an emphasis on the moral (as compared to the competence) test implications, participants were more likely to control their negative bias against Muslim women.

In this dissertation research, no specific norms were made salient when the moral implications of the task were emphasized, nor were participants explicitly informed on how they might avoid displaying bias while working on the IAT. In the moral motivation condition, participants only read that the test could give an indication of the value they attached to egalitarianism vs. discrimination. Furthermore, participants' performance was assessed when they performed the test in private and anonymously. The findings obtained with this procedure extend previous research as they make it possible to exclude a number of alternative explanations relating to self-presentation and displays of socially desirable response patterns. Thus, the data reported here reveal that people act upon their own moral values, presumably because this is important for how they see themselves.

### **Part II: Moral Motivation is Affected by the Social Context**

In the second part of this dissertation, I examined whether it is also important for people to be perceived as moral by others. To examine this I introduced a procedure that led participants to believe that their performance on the IAT was monitored by another individual present in the lab. My findings reveal that people are particularly motivated to act according to their moral values in the presence of people who belong to the same group as they do (i.e., ingroup members). In my research, these groups were created according to very minimal criteria (i.e., a minimal group paradigm; Tajfel, 1970). Before participants started with the IAT, they completed a questionnaire that ostensibly assessed their personality type. After that, they were told that they were coupled with their evaluator based on both their questionnaire scores. It was explained that when the evaluator was assessed to have the same personality type as the participant, they shared this particular group membership. In other words, this made them ingroup members. When the evaluator was assessed to have another personality type, s/he differed from the participant. This distinction in their personality types thus made the evaluator an outgroup member. This type of paradigm allowed me to exclude the possibility that alternative concerns (such as prior liking, familiarity or value similarity) might induce participants' responses to different evaluators. Thus, I was able to establish that people are more motivated to act morally in front of others who are relevant to the self. In real life, these self-relevant others might include people with the same

nationality, gender, religion, or occupation. The current findings thus extend previous research that demonstrated that people explicitly report the importance of being seen as a moral ingroup member (Leach, Ellemers, & Barreto, 2007) by showing that they actually are more likely to act in accordance with their moral values when their behavior is monitored by an ingroup member.

One could wonder how meaningful the situation created in this experiment is, as a minimal group paradigm is unrelated to the intergroup associations examined with the IAT. That is, in the IAT participants are asked to make associations between non-Muslims and Muslims and pictures of positive and negative scenes. Group memberships based on personality type may thus be seen as irrelevant to the task. However, exactly because of the use of a minimal group paradigm, I was able to reveal the importance of being perceived as moral for people's social identity. Introducing two experimentally created groups, which had no meaning or known moral values outside of the laboratory, was sufficient to increase people's motivation to appear moral towards an ingroup rather than an outgroup member. This finding can thus not be attributed to factors other than the categorization allegedly based on personality types introduced in the experiment.

Nevertheless, the importance of being perceived as moral by self-relevant others may go beyond a shared minimal group membership. In fact, introducing a group membership that does interfere with the intergroup associations made in the IAT may reveal additional motivations to adhere to moral norms. In Chapter 5 I accordingly established that being perceived as unprejudiced is even more important when a representative of the social target group is present. That is, when participants thought that their performance was monitored by a Muslim woman, they inhibited their bias against Muslims to an even greater extent than in the presence of a minimal ingroup member. Although this finding is consistent with previous research on intergroup bias (e.g., Lowery, Hardin, & Sinclair, 2001; Richeson & Ambady, 2003), the current results extend this research in an important way. The way my study was set up allowed me to show that the moral implications of one's behavior can be emphasized in many different ways – that all can be effective. The results of Study 5.1 show that simply mentioning the moral implications of the task affected people's implicit prejudice in a similar vein as did

the actual presence of a Muslim evaluator. That is, my research demonstrated that participants inhibited their bias to a similar extent when they were being monitored by a non-Muslim woman during a task of which the moral implications were emphasized as when they were being monitored by a Muslim woman. Taken together, these findings thus reveal that signaling the moral implications of one's performance can instigate moral behavior to a similar extent as explicitly confronting people with others who depend on them for moral treatment (i.e., Muslims).

## **The Underlying Processes**

### **Parts I and II**

Apart from showing behavioral effects of emphasizing the moral implications of one's behavior, I also examined the cognitive processes underlying people's motivation to be and to appear moral. More specifically, by applying measures borrowed from the field of neuroscience, I was able to show how focusing on people's morality changes their attention to ingroup and outgroup members, as well as the degree to which they monitor their own moral behavior.

#### **Moral motivation changes people's focus of attention.**

In Studies 2.2, 3, and 4.2, I examined brain activation associated with the motivation to be moral using event-related brain potentials derived from EEG (i.e., ERPs, derived from activation recorded at the scalp) and functional MRI (i.e., to localize activation in the brain) while they were performing moral behavior. Results showed that emphasizing the moral implications of people's behavior causes them to increase their attention towards the faces of the different group members presented in the IAT. People thus attended more to the difference between ingroup and outgroup members when they were motivated to approach this task in a moral way compared to when they were concerned with being competent at the task. At first sight this increased attention to group membership may seem to contradict moral intentions. That is, performing in line with moral values –not revealing intergroup bias– might also be expected to result in less differentiation between groups evident in increased similarity of cognitive responses when looking at members of ingroups and outgroups. Nevertheless, while participants were more inclined to attend to the group membership of the target stimuli under moral task



instructions, we found in Study 2.2 that people were more able to respond in an unbiased way. This combination of effects seems to suggest that the increased social categorization of ingroup as distinct from outgroup members was needed in order for participants to inhibit their bias against the outgroup and thus to adhere to their moral values. This explanation is in line with the notion that in order to deal with one's upcoming thoughts, these must first be recognized and accepted (Wegner, 2011). Likewise, in order to suppress the tendency to reveal bias, one must first acknowledge the difference between group members.

The investigation of the cognitive processes underlying moral motivation also extend our findings on the behavioral measures revealing the importance people attach to being perceived as moral by their ingroup members in particular. That is, complementing the behavioral effects observed in Study 4.1, the results of Study 4.2 revealed that participants' increased cognitive attention to the ingroup and outgroup faces when the implications of the test were formulated in terms of their moral values, only emerged when they were evaluated by someone of their own (minimal) ingroup, and not when they were being monitored by a member of another (out)group. In other words, the adjusted cognitive approach towards the task –arguably to make it possible to adhere to moral group norms– was especially apparent in an intragroup context.

#### **Moral motivation enhances response-monitoring.**

Besides the increased perceptual attention to the difference between faces of ingroup and outgroup members, participants to whom the moral (rather than the competence) implications of the task were emphasized also showed enhanced error-monitoring. That is, when participants were motivated to show their morality, they paid more (automatic) attention to their *responses* than when they wanted to show their competence. Consistent with previous ERP findings (e.g., Gehring, Goss, Coles, Meyer, & Donchin, 1993), the error-related negativity (ERN) modulation was evident when participants made incorrect (as compared to correct) responses. Importantly, this enhanced response to errors was greater when the moral rather than competence implications of the task were emphasized. Previous research has revealed that increases in the ERN are associated with how important a good task performance is to people, which is indicated by the extent to which

they care about making errors on the task (Hajcak, Moser, Yeung, & Simons, 2005). The enhanced ERN modulation in case of an emphasis on one's moral values thus implies that people are more concerned about making mistakes when the task supposedly indicates their morality than when it 'merely' indicates their competence. Additionally, these findings suggest that the motivation to be moral can in part be explained by an increased concern about not appearing moral. In comparison, the prospect of appearing incompetent seems to be less distressing.

Importantly, the effect concerning participants' error-monitoring also depended on the social context. Results of Study 4.2 showed that the emphasis on moral implications in combination with being monitored by an ingroup member increased response-monitoring on both incorrect *as well as* correct responses. Thus, when people are evaluated by another ingroup member, they show a general increase of attention towards their own moral behavior. The ERN findings in Part I thus reveal that people are primarily concerned with making mistakes that can be perceived as an indication of immoral behavior. In addition, the results in Part II show that when people show their moral behavior to their fellow group members, it seems equally important to detect any mistakes as it is to monitor their correct responses.

### **Moral motivation increases detection of task-relevant characteristics.**

The emphasis on the moral implications of the task and being monitored by an ingroup member also affected participants' detection of the different types of trials in the IAT. The IAT consists of incongruent and congruent trials. As participants who took part in the research described in this dissertation were non-Muslim, the congruent trials consisted of associating faces of non-Muslim women (i.e., ingroup members) with pictures of positive scenes, and faces of Muslim women (i.e., outgroup members) with pictures of negative scenes. By contrast, the incongruent trials consisted of associating outgroup members with positive pictures and ingroup members with negative pictures. Previous ERP research has shown that the detection of the difference between such congruent versus incongruent trials (i.e., 'conflict-monitoring') is visible in the N450 modulation, which is typically larger for incongruent than congruent IAT-trials (e.g., Williams & Themanson, 2011). Results of Studies 2.2 and 4.2 showed that the N450

modulation was increased in case of an emphasis on morality and when an ingroup member was evaluating participants' performance. The detection of the different types of IAT trials was thus enhanced under these circumstances. A possible explanation for this finding may be that the moral implications of the test, and the presence of an ingroup member, made the meaning of the difference between congruent versus incongruent trials more evident. It suggests that participants may have understood that the ease with which they responded on congruent as compared to incongruent trials was related to possible signs of prejudice. These participants may have realized that the relatively easy part of the task consisted of associating the outgroup with negativity and the ingroup with positivity. And that the relatively difficult part meant associating the outgroup with positivity and the ingroup with negativity. In contrast, participants who thought the task was indicative of their competence may only have noticed the difference in the level of difficulty between the two types of trials, without taking a notion of the social meaning of the associations they were asked to make.

Overall, the findings of Parts I and II are important as they extend prior research that used self-reports (e.g., Leach et al., 2007; Ellemers et al., 2008) as well as our own research showing actual moral behavior on an IAT to examine the importance of being moral. By incorporating the examination of unconscious cognitive processes with ERP measures, the current findings reveal *how* people's motivation to be (perceived as) moral leads to more moral behavior. Results concerning the underlying cognitive processes reveal that moral concerns affect how people perform the task and to what kind of aspects they pay attention during the task (i.e. "Is this person a Muslim or non-Muslim?"; "Is this particular trial more or less difficult?" and "Am I succeeding in being unbiased?"), affecting their actual moral behavior (in this case implicit bias against Muslims).

### **Part III: People Show a Positivity Bias Concerning Their Own Morality**

Overall, the behavioral, ERP and fMRI results of the first two parts of this dissertation indicated that emphasizing the moral implications of one's behavior (either while being evaluated by an ingroup member or in private) causes people to become more vigilant during their performance on a test of implicit prejudice. The findings also seem to suggest that adherence to moral norms is equally important as

it is to avoid committing moral transgressions. A possible explanation could be that the motivation to be moral is accompanied by a fear to appear immoral, whereas the possibility of appearing incompetent may be less distressing. That is, in our mind even competent people may sometimes do incompetent things, but people who do something immoral once are unlikely to be seen as moral persons. In Part III of this dissertation, I therefore examined how much people care about whether they succeed or fail in behaving according to their moral values – compared to how much they care about their success or failure in the competence domain.

In Chapter 6, I assessed people's affective and cognitive responses after and while they received information about their own moral and competent behavior. Participants first performed a task (the IAT), but in contrast to the procedure in previous chapters, this task was said to be indicative of their moral values *as well as* their competence. Thereafter, they were either informed that they had performed above (positive feedback) or below (negative feedback) average on the moral and competence dimensions of the task. This allowed me to directly compare how positive versus negative feedback concerning one's own moral and competent behavior impacted upon people's state of mind and emotional well-being.

Results of Study 6.1 revealed that people feel bad when they are confronted with information indicating that they are not that moral as compared to others. Such information causes increased levels of physical arousal (measured using skin conductance responses) and people also report to experience more intense negative emotions. Crucial to my predictions, receiving information that one is less moral than others makes people feel worse than receiving information indicating that they are less competent than others. These findings thus confirm the notion that people care more about whether they succeed in behaving according to their moral values rather than behaving competently.

Additionally, results of the fMRI study in Chapter 6 seem to suggest that when people receive positive information indicating that they are more moral compared to others, they perceive this information as highly relevant to their self-concept. Previous neuroimaging studies showed that activation in the (ventral) medial part of the prefrontal cortex (vMPFC) is associated with ascribing personal characteristics or behaviors to the self (e.g., Van der Meer, Costafreda, Aleman, &

David, 2010). In line with the notion that people want to be moral, I thus examined whether viewing information indicative of one's own moral behavior is associated with activation in the vMPFC. Indeed, results of Study 6.2 showed that activation in the vMPFC was greater when participants viewed feedback about their moral behavior as compared to their competent behavior. Interestingly, this was only the case when this feedback consisted of positive information. The results of Chapter 6 thus reveal that people seem to perceive positive indicators of their moral behavior as particularly self-relevant. A tentative explanation for this finding could be that participants in this study protected themselves from negative feedback by processing it as relatively less self-relevant. This is in line with my observation that the confrontation with negative indicators of one's own morality has a highly negative impact upon people's emotional well-being. Hence, discarding such information as less self-relevant might be part of a self-protective strategy to cope with such threatening information. Taken together, the findings thus confirm how much people care about succeeding in behaving in line with their moral values, and how they respond to information that may indicate this.

### **The Added Value of Different Research Methods**

In this dissertation, I addressed three research questions related to people's motivation to be (perceived as) a moral individual and group member. A primary aim of the dissertation was to examine the underlying processes associated with this motivation. I thus combined behavioral observations with psychophysiological and neuroscientific research tools throughout the empirical chapters to go beyond observing *what* people do, and examined *how* and *why* they do this in terms of specific underlying processes.

The behavioral task used in the empirical chapters provided reaction times and error rates. It showed us that people inhibit their bias against Muslim women by slowing down their responses on prejudice-congruent trials. This measure thus revealed *what* people do, but it remains unclear *how* they are able to do this. Likewise, self-report measures are often administered after a particular behavior is displayed. Such measures rely upon the ability and willingness of research participants to report on their psychological state while performing the task, and are sensitive to social desirability – which obviously is a significant factor in

research concerning the motivation to appear moral. This is why it was important for me to assess the psychophysiological and neuroscientific measures *online*, that is, while participants were actually performing the task. The neural and physiological reactions I assessed occur unconsciously and are less sensitive to the intention to respond in socially desirable ways.

Using ERPs, I was able to disentangle different cognitive processes associated with the control of prejudice. In this way, I revealed three different mechanisms that help participants inhibit their (behavioral) bias against Muslims. They did this: (a) By (unconsciously) increasing their perceptual attention to categorize target faces as Muslim versus non-Muslim women; (b) by distinguishing between prejudice-congruent and –incongruent trials; and (c) by monitoring their responses during the task. Recording skin conductance responses allowed me to reveal that receiving information about people’s own moral behavior causes instant automatic arousal that is different from how they respond to information concerning their competence. These findings thus underscored participants’ explicit reports of their negative affective states. Furthermore, based on fMRI-results –and particularly the knowledge of the functional properties of activation in the ventral medial part of the prefrontal cortex– I have suggested that people perceive positive information indicating their morality as particularly relevant to their self-concept.

To give a concrete example of the added value of the different research methods combined in the current dissertation, let’s consider the findings in Chapters 2 and 4. Here, I discovered that an emphasis on the moral implications of one’s behavior affects people’s approach towards a task. Using a behavioral measure of implicit prejudice, I showed that non-Muslim participants inhibited their negative bias against Muslims when they were told that the test could assess their moral values concerning egalitarianism (as compared to how competent they are; Studies 2.1 and 4.1). The weaker negative bias when the moral test implications were stressed was caused by a smaller difference in response times between incongruent and congruent trials. This means that when morality was emphasized, non-Muslim participants responded equally slowly to congruent trials (associating non-Muslims with positivity and Muslims with negativity), as to incongruent trials (associating non-Muslims with negativity and Muslims with positivity). In this

sense, people thus made less of a distinction between their associations with ingroup and outgroup targets, and this is what resulted in the reduction of bias. However, interestingly, examination of brain activation during task performance revealed a significant difference between viewing pictures of ingroup and outgroup targets. That is, ERP modulations associated with differentiating between viewing in- and outgroup targets were increased rather than decreased (Studies 2.2 and 4.2). Additionally, activation in the occipital face area was greater for viewing faces of ingroup compared to outgroup targets when morality was emphasized (Study 3).

At first glance, the behavioral and neuroscientific research findings thus seem to be contradictory. On a behavioral level, emphasizing morality resulted in more *equal* responses to members of different groups, whereas emphasizing morality actually increased *differentiation* between groups at the neural level. However, it is important to understand that both measures assessed different cognitive processes which occur at different stages in the process. That is, behavioral bias was estimated based on reaction times and the accuracy of responses on all trials within the task. This includes trials with pictures of faces of in- and outgroup targets, and on trials with pictures of positive and negative scenes. By contrast, perceptual attention was assessed from early visual processing of faces alone, irrespective of the response given on these types of trials. This may imply that the salience of morality *increased* people's perceptual *attention* towards the faces of in- and outgroup members, and this is what enabled participants to *behaviorally* respond with *decreased* bias, in line with their moral values. It also suggests that participants actually attended differently to specific task stimuli when its moral implications were emphasized, rather than merely correcting their behavioral responses to these stimuli. This combination of observations thus suggests that the adjustment in participants' behavior may at least in part depend on early cognitive processes that are crucial for preparing these responses.

Another example of the added value of the new approach I followed in the current dissertation concerns the examination of the behavioral IAT effect. Across the different studies reported in Parts I and II, I found the same effect of emphasizing the moral implications of participants' IAT performance. Participants to whom the moral implications of the task were emphasized (and whose

performance was monitored by an ingroup member) showed a relatively weak negative bias against Muslims. Interpreting this finding based only on the strength of the bias does however not inform us about how this reduction in bias is achieved. For example, the emphasis on morality may have caused participants to develop stronger positive associations with Muslims. On the other hand, it could also have helped them to control their negative associations with the Muslim targets. Either way this might have increased perceived equality between the two target groups, resulting in the smaller negative bias against Muslims that was found. Nevertheless, I set out to examine which process actually resulted in these behavioral findings.

Previous research concerning the malleability of implicit bias can be seen to represent two distinct approaches. There are studies (such as the ones described in this dissertation) that examine what kind of motives or contextual factors affect displays of prejudice (i.e., the dependent measure of social bias such as the IAT effect). There are also studies that focus on how people's performance on such measures of prejudice can be influenced (e.g., Fazio & Olson, 2003; Olson & Fazio, 2003). In this second type of research different models have been introduced to examine responses on reaction times and error rates to disentangle the processes underlying automatic evaluations and control. Examples are the process-dissociation model (e.g., Jacoby, 2001; Payne, 2001); the Quad-model (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005); the diffusion-model analysis (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007); and the ReAL model (Meissner & Rothermund, 2013). In my research, I did not follow any of these particular models, but adopted a more general strategy to examine the underlying processes associated with the reduced bias in case of an emphasis on moral values. I more closely examined participants' response times on *correct* congruent and incongruent IAT trials, to be able to distinguish between two different routes towards bias reduction.

In theory, bias measured using an IAT can be diminished in two ways. Either by reducing response times on incongruent trials, or by increasing response times on congruent trials. The first strategy implies that participants try to respond faster when they are asked to associate outgroup members with positive attributes and



ingroup members with negative attributes. The second strategy implies that participants slow down their responses when they are asked to associate outgroup members with negative attributes and ingroup members with positive attributes. As described in each of the relevant chapters, the current results showed that the emphasis on morality (and being evaluated by an ingroup member) caused participants primarily to *slow down* their responses on prejudice-*congruent* trials. This suggests that stressing the moral implications of their performance made the meaning of congruent trials –as potentially revealing biased associations– more salient. That is, under moral task instructions participants were more likely to realize that congruent trials would reflect the ease with which they could associate Muslims with negative attributes and non-Muslims with positive attributes. My approach to examine how the emphasis on moral values affects people’s bias towards Muslims thus revealed that this may have led them to slow down and overthink these prepotent responses, to be able to act in line with (self-relevant) moral values.

### **The Challenges of Different Research Methods**

In my research, I set out to combine procedures and measures that had been developed in different research traditions, to examine distinct research questions. Combining different approaches in this way certainly had an added value for my research and the conclusions I was able to draw. Nevertheless, I also had to face several complications relating to adjustments I had to make to experimental designs and standard procedures, to adapt the IAT for use of different neuroscientific research methods.

In Study 2.2, I had a clear hypothesis about how the ERN modulation would be affected by the emphasis on the moral implications of the task. But in order to reliably estimate the ERN, a sufficient number of errors is needed. I thus doubled the amount of trials in the IAT in order to allow participants to reveal more mistakes during their task performance. Although prolonging the IAT did result in the intended increase in errors, it also caused a learning effect: After so many trials, regardless of condition, all participants responded in the same way to all types of IAT trials. This adaptation of the task to enable examination of the ERN modulation thus reduced the difference in performance depending on whether

moral or competence task implications had been emphasized. As a result, the behavioral effect I had found in Study 2.1 was less clearly visible in Study 2.2.

In Chapter 3, I faced a similar problem, when I did not find an effect of emphasizing moral test implications on the behavioral data at all. In this fMRI study, I used an event-related design to be able to detect brain activation associated with the presentation of the different types of stimuli. To be able to separate responses to different trials, this design requires that there is a certain waiting period in between each of the trials. The time delay between trials, required to reliably assess fMRI responses, slowed down the overall pace of the IAT, and may have helped participants to prepare for and focus their attention for each upcoming trial. As a consequence, when using this procedure, the response times of participants who read the moral implications of the test were similar to the response times of participants who read the implications concerning their competence. This aspect of the task procedure may explain why I was unable to demonstrate the previously presented behavioral effects of the emphasis on one's moral values, in this particular study.

Finally, the behavioral effect of reduced bias in case of the morality frame in combination with evaluation by an ingroup member (observed in Study 4.1) did not emerge in (ERP) Study 4.2 in which I examined the effects of morality framing and presence of ingroup versus outgroup members. In retrospect, this may be attributed to the limited response window we offered to participants. As was the case in Study 2.2, I adapted the IAT procedure in Study 4.2, because I needed enough erroneous responses to reliably estimate the ERN. A pilot study had however uncovered that participants responded more accurately as well as faster when their performance was being monitored. Thus, in addition to doubling the number of trials like I did in Study 2.2, in Study 4.2 I also tried to induce participants to make a sufficient number of errors by reducing the time available to respond on each trial. In Studies 2.1, 2.2, and 4.1, the decrease in behavioral bias against Muslims was associated with participants' slowed down responses on congruent trials. Slowing down was however no longer possible in Study 4.2 because of the limited response window. This might explain why no evidence of reduced behavioral bias was found here. Nevertheless, the ERP measures

confirmed that the underlying cognitive processes were affected when the moral implications of the task were emphasized and when participants were monitored by an ingroup member.

Unfortunately, such difficulties are inherent to the choice of combining different research methods –while using the same behavioral paradigm–, to obtain triangular evidence as a way to examine complex psychological questions (see also Scheepers, Ellemers, & Derks, 2013). However, importantly, the fact that these adaptations had to be made and affected the results also extended current insights in the processes underlying the influence of (moral) motivation on IAT performance. For example, in Study 5.2, I also extended the number of trials included in the IAT. This time, I examined whether increased exposure to an apparent outgroup member (i.e., a woman with a headscarf) who was presented as a partial ingroup member on another dimension (i.e., in terms of her personality type) might increase positive associations with Muslim women. As in Study 2.2 – where the IAT effect was extinguished over time– the prolonged IAT caused a learning effect once again. But this time extending the number of IAT trials enabled participants to *develop new associations*, by learning to combine positive stimuli with the outgroup target. The evidence that it is possible for participants to do this is important beyond its methodological implications, as it offers scope for developing very practical and concrete strategies that may help reduce the emergence of implicit negative biases by learning to make new associations (see also Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000).

### **Extending Previous Literature**

The research presented in this dissertation extends existing insights in many ways. Importantly, this research is the first to show that morality not only induces people to *say* that they want to behave in a certain way, but that it actually motivates them to change their *behavior*. Extending previous research that focused on people’s explicitly reported moral motivation and stated preferences in hypothetical moral dilemmas, I was able to reveal that people adjust their implicit behavior (i.e., their IAT performance) in line with their moral values when the moral implications of that task are made salient.

The findings reported in the current dissertation also have important implications for neuroscientific research. Previous research has examined the brain regions involved in the ability to behave in line with moral standards by studying patients with brain lesions who –as a result– exhibit immoral behavior or psychopathological characteristics (e.g., Anderson, Bechara, Damasio, Tranel, & Damasio, 1999; for a review see also Moll, Zahn, De Oliveira-Souza, Krueger, & Grafman, 2005). In other cognitive research there has been a focus on moral decision making. There, it is examined which parts of the brain people need (i.e., need to be activated) to *consider* what one would do in a hypothetical situation. The findings in this dissertation thus extend those insights by revealing the cognitive processes involved when healthy participants *behave* (i.e., perform) in line with their own moral values. Moreover, by studying the brain regions involved in the motivation to appear unprejudiced in a situation that resembles common interactions (when first viewing faces of people representing different social groups), I was able to examine a kind of moral motivation that is part of social interactions and thus of everyday life.

### **Societal Implications**

Besides the theoretical implications of the current dissertation, the findings presented here also have some important practical implications. In Part I, I revealed that emphasizing moral implications of people's behavior caused them to inhibit their behavioral prejudice towards outgroup targets. This could imply that in real life settings, people may also adjust or control their behavioral or verbal expressions of prejudice when they are made aware of what such expressions might say about their own moral values. Consider for example a situation in which an employer rejects an applicant, merely because she indicates she wants to wear a headscarf at work. In this situation, the employer will probably only be aware of the consequences for the applicant rather than thinking about what the rejection of such applicants may reveal about himself and the company more generally. Awareness of the implications of his behavior in terms of his own morality and what it says about his values regarding equality and intercultural respect may make him more careful to ensure equal treatment in future interactions.

The findings reported in Part II of this dissertation also reveal that people care whether they are perceived as moral by *others*, especially by members of their own group. Motivating people to inhibit their prejudice towards outgroup targets by emphasizing moral values instead of competence, will thus be particularly effective within a group context. This finding speaks to debates about how to best promote diversity policies in work settings. In the literature, it has been suggested that, rather than emphasizing that diversity is the *right* thing to do, organizations should emphasize the ‘business case for diversity’. In essence, this is a focus on *competence* that may persuade the executive board of a company to work towards a more diverse organization, and promotes ethnic and gender diversity in order to improve the organization’s profit and success. In addition, the ‘business case for diversity’ is proposed to increase motivation and efficiency among employees (e.g., European Commission, 2005; Robinson & Dechant, 1997). However, in terms of diversity climate within the company, based on the results of this dissertation, I question the effectiveness of this measure in motivating employees to embrace diversity and treat colleagues from other ethnic backgrounds with respect. Instead, a better way to achieve this might be by emphasize that striving for a diverse organization is the *moral* thing to do.

Presenting policies in terms of moral principles, to motivate members to act accordingly may be a strategy that can actually be adopted by any kind of company, department, or team. Consider for instance organizations in the financial sector. Here, norms and performance targets also tend to be presented in terms of competence. In order to make a profit, close a successful deal, or attract new clients one should first and foremost be clever and skilled. Although this may sound intuitively convincing, my findings imply that it might be even more motivating for employees to be part of and work for an organization that emphasizes its moral character, for instance by focusing on fair treatment of employees, or showing honesty towards clients. Indeed, there is some correlational evidence in line with this reasoning, documenting that perceptions of organizational morality relate to employee satisfaction and work commitment (Ellemers, Kingma, Van de Burgt, & Barreto, 2011).

The notion that evaluations by other ingroup members are particularly effective in helping people to display moral behavior is also relevant in the context of the financial sector. In the Netherlands, organizations in this sector are being supervised and controlled by an external agency, the Authority of Financial Markets. This is an independent institution assigned to check and sanction the proper business conduct of financial markets, accountants or other financial service providers. However, an important consequence of this independent supervision is that evaluations of ethical business conduct are made by an *external* source. In the terminology used in my research, this would represent an outgroup judgment. That is, a judgment from a group that people tend to consider less self-relevant, which may for this reason alone be less effective in influencing their moral behavior. In view of the findings reported in this dissertation, it may be questioned whether supervision from such a source provides an optimal way to guide adherence to moral standards. If the goal is to improve morality in the financial sector, it might be more effective when moral norms are emphasized *within* a company and by its *own* board. Having moral business conduct as a core company value, is more likely to stimulate employees to perform their work in line with ethics guidelines.

Although emphasizing moral rather than competence norms may be particularly effective within one's own group, the findings of Part II of this dissertation also show that concerning people's control of prejudice, moral behavior can be influenced when this is evaluated an outgroup member. That is, people will generally be inclined to inhibit the expression of their negative bias when they are being monitored by a member of the group that is the target of such bias. This implies that diversity in a setting where people cooperate or evaluate one another may prevent displays of prejudice and discrimination. For example, having a Muslim employee as a member of an evaluation committee and who will thus observe the decision-making process concerning candidates for the job, may thus help the committee to create equal opportunities for Muslim as well as non-Muslim applicants. Likewise, having women present in the board of directors of a company may help others control gender bias when considering applications for high-status positions.

Importantly, my research also revealed that emphasizing the moral implications of people's behavior can be just as effective in reducing prejudice as is the presence of a representative of the group that is the target of prejudice. This is important for contexts in which intergroup contact is not feasible. This is the case for instance, when employees do not yet have any colleagues with a different ethnic background or religion, but might be induced in this way to be more open and welcoming to such a colleague. Likewise, emphasizing moral implications of being unbiased might be of benefit in the integration of newcomers in neighborhoods that primarily consist of people from the same social class or ethnicity. Within these contexts, people may be motivated to control their prejudice when this is emphasized as the right or moral thing to do, giving the new colleague or neighbor a fair chance to reveal their personal qualities rather than relying on biased expectations. Standard communications regarding company policy or national campaigns to encourage equal treatment of minority members tend to focus on the negative implications for the targets of prejudice, as a way to prevent people from expressing bias. My research suggests that there is likely to be added value in communicating about moral values and equality goals of the perpetrators, as a way to help diminish prejudice.

The research reported here not only elucidates how people adapt their moral behavior, it also reveals some very concrete and practical ways in which moral behavior can be stimulated. However, in real life, we have to take into account that even with the best of intentions people may sometimes deviate from what is considered moral, or be unable to always live up to their moral standards. The findings in Part III of this dissertation reveal how people are affected when confronted with their own moral slips. Because of the motivation to do what is morally right, confronting people with their moral failures has a negative impact upon their emotional well-being. If this negative response is sufficiently severe, it is likely to induce feelings of inadequacy and stress, which people are likely to cope with through denial or motivational withdrawal. Indeed, some of the fMRI evidence seems to suggest that negative moral information tends to be seen as less relevant to the self, even if skin conductance responses and self-reports indicate that receiving this type of information clearly has an emotional impact. Thus,

emphasizing their moral failures may not be the best way to motivate people to change or improve their behavior. Importantly however, people also seem to be especially attentive to positive information about their moral behavior which they seem to perceive as particularly relevant for their self-concept. This is relevant for instance to leaders who have to monitor and sanction the behavior of their subordinates. The natural tendency might be to confront an employee with moral failures, such as unethical decision making, as a way to prevent similar behavior in the future. However, due to the negative emotional impact this has, this might not be the best way to achieve behavioral change. Instead, it might be more effective to encourage the employee to succeed in their motivation to be moral by emphasizing moral achievements, or praising them for compliance with moral norms or company values while doing their job.

Thus far, I have mainly focused on the practical implications of the current findings in business settings. However, in principle, emphasizing the moral implications of people's performance can also be effective in stimulating moral behavior in other contexts. For example, similar mechanisms might be effective in sectors such as sports where moral behavior may be enhanced by emphasizing the importance of fair play and proper competition instead of focusing on winning outcomes alone.

### **Limitations and Directions for Future Research**

The results of the empirical chapters outlined in Parts I, II, and III offer new insights in people's moral motivation. However, there are also some limitations in the studies described in Chapter 2-5 that need to be addressed and which may provide directions for future research.

A possible point of critique concerning the current dissertation is the repeated use of the implicit association test (IAT). In fact, this was the only (implicit) measure used to examine moral behavior in this thesis. In this research, the IAT was chosen because this measure lends itself rather well for framing its implications in terms of morality *and* in terms of competence. Also, the use of multiple trials makes it a measure of moral behavior that is also viable for the examination of cognitive responses, which requires repeated behavioral displays to achieve a reliable assessment of underlying processes. Moreover, the consistent use of the



IAT made it possible to compare and combine the related results of different scientific measures of cognitive processes and brain activation. Nevertheless, in future research other experimental paradigms might be developed to examine the importance of revealing one's motivation to be moral over one's motivation to be competent. For example, it would be interesting to examine whether people's own motivation to be moral rather than competent would also affect behavior in more economic type of situations, such as in bargaining games, where there is a clear trade-off between moral concerns (e.g., fairness, trust) and competence concerns (e.g., outcomes). That is, extending the current research as well as studies that examined the effects of knowledge about the moral character of the *other* player on the behavioral choices in such games (e.g., Delgado, Frank, & Phelps, 2005; Frank, Gilovich, & Regan, 1993), it could be tested whether morality is a stronger motivator than competence for people's own choices in situations where moral behavior may go at the expense of individual outcomes and ingroup norms trump individual gains.

Additionally, I have examined the importance of being perceived as moral by others, by introducing intra- and intergroup contexts. The examination of individual differences in the motivation to adhere to specific moral norms was not the focus of the current research. Nevertheless, previous research concerning prejudice control and automatic evaluative associations has revealed that such individual factors do explain differences in the regulation of social bias. For example, people can be internally and/or externally motivated to respond without prejudice on particular assessments (Plant & Devine, 1998; Amodio, Harmon-Jones, Devine, 2003; Amodio, Kubota, Harmon-Jones, & Devine, 2006). In some studies, participants are even preselected based on a measure of this motivation. For example, Amodio et al. (2006) recruited research participants who were previously found to score high on the internal motivation scale and low on the external motivation scale, to compare their responses with those of people who score high on both scales. In some of the chapters in this dissertation, an assessment of internal/external motivation to avoid prejudice was included as an additional background measure. In my research, participants generally showed internal instead of external motivation to appear unprejudiced. This is consistent

with the notion that I examined the motivation to be moral as a self-relevant goal. Future research might seek out research participants that are primarily externally motivated to appear unprejudiced. This might make it possible to examine for instance whether such individuals are less sensitive to feedback concerning their own morality, but might be more responsive to moral evaluations by others. Now that I have established these different concerns as relevant to the adaptation of moral behavior, it might be of interest to specify whether certain groups of individuals might be more open to certain types of moral interventions than others.

### **Conclusion**

Using different scientific research methods, the findings in this dissertation reveal that (1) people tend to act in ways that are considered moral; (2) it is important for people to be perceived as moral by self-relevant others; and (3) that people care about succeeding in behaving according to their moral values. The findings extend previous research by observing and measuring people's actual behavior. Furthermore, automatic brain and physiological responses revealed how people respond to and initiate behavior in order to adhere to their moral values.