

The role of emotions in moral judgments

Annemieke Krijger

Erasmus University, Rotterdam

Abstract

This study investigated whether the emotions contempt, anger, and disgust were specifically related to a moral judgment in the category community, autonomy, and divinity respectively according to the CAD-hypothesis (Rozin et al., 1999). In Experiment 1 an emotion (anger, neutral, or disgust) was induced suboptimally or optimally by showing a face. Then a sentence appeared and the participants had to judge the described situation (good or bad) as quickly as possible. In Experiment 2 participants had to read morally bad sentences and had to choose the face (emotion) that the described situation would evoke. No evidence for the CAD-hypothesis was found. The results of both experiments did not support a specific relation between category and emotion.

Introduction

Why is it wrong to gossip about a colleague? The answer seems obvious: because you can hurt someone psychologically. And why should you not practise kissing with your brother or sister? If we have to decide whether this is morally good or bad, most people will conclude that it is not done. But why is it not done? It is difficult to find convincing rational arguments why this is 'bad'. The purpose of this article was to investigate the proposal that moral judgments are not primarily guided by rational argument but rather by the emotions we experience when exposed to moral situations.

Before considering how moral judgments are made the grounding of morality in the brain is taken into account. How do we know something is good or bad? Are prior moral situations stored in abstract, language like, non-perceptual codes (amodal theory) or perceptual, analogue symbols (modal theory)? The development of the computer and particularly the programming of language formed an important source of inspiration for the amodal theories. According to Hume on the other hand morality is grounded (embodied) in feelings of moral approval or disapproval (Hume, 1751/1983). John Locke (1632-1704) and still much earlier Aristotle (4th BC) and Epicurus (4th BC) already thought of higher cognitions and representations as perceptual and based on experiences.

In cognitive science the same line is extended. Lakoff and Johnson (1999) suppose the mind is embodied and abstract moral concepts are structured metaphorically. Moral metaphors are formed through experience of well-being, especially physical well-being. This idea corresponds largely with the perceptual symbol theory as described by Barsalou (1999). This theory implies that perceptual states, at the moment of experience and if attention is directed to that experience, are stored in analogue, modal symbols instead of amodal symbols. These perceptual memories will act

symbolically and will be activated in the appropriate situation. This is called the reenactment process (Barsalou, 1999; Barsalou, Niedenthal, Barbey, & Ruppert, 2002). This process does not need to be necessarily conscious. Most of the embodiment effects result from automatic processes. Two constructs are important considering reenactment: simulators and simulations. The simulator is the place where the simulation can be constructed. In memory an infinite number of simulators can be developed for all different kinds of knowledge. The states of several modalities are stored in this simulator like information about perceptual, motor, and introspective states. Applied to abstract constructs the simulators contain more situational and introspective state information than the simulators of concrete concepts (Barsalou, 1999; Barsalou et al., 2002). Prinz (2003) applied this theory to moral concepts as a single class of abstract concepts. According to Prinz (2003) moral truths are evaluations. Moral judgments are grounded in observations of ourselves. If something is morally bad an aversive response to it is perceived. The perception of change in one's bodily state is an emotion. Thus the concept of morality is perceptually grounded (embodied) by experiencing an emotion.

The debate whether moral judgments are based on rationality as described by Immanuel Kant (Sullivan, 1989) or emotionality (Hume, 1751/1983) has a long history (Prinz, 2003; Rozin, Lowery, Imada, & Haidt, 1999). With regard to the rationalistic view cognitions and reasoning make up a moral judgment. The six stage theory of moral development by Kohlberg (1969) is largely cognitive and inspired by the cognitive development theory as described by Piaget (1965). The theory of moral development states that moral judgments are based on moral reasoning and moral reasoning is supposed to be a conscious, language related process. A main shortcoming in Kohlberg's theory is the ignorance of the importance of emotions related to moral judgments. Gibbs (2003) offered a kind of compromise between the two extremes rationality and emotionality. His solution was that moral motivation is neither 'affective primacy' (Hoffman's position) nor 'cognitive primacy' (Kohlberg's position) but 'coprimacy' (both an emotional and a rational response as primary motivations). Hume (1751/1983), a British empiricist, also stated that a moral judgment is based on both reason and sentiment. But the experience of a sentiment (approval or disapproval) was the most important part in making up a moral judgment. Intellectual faculties are needed for 'assistance'.

The empiricist view was the main view until the cognitive revolution. The rationalistic view was dominant since the emergence of cognitive psychology in the 1950s (Barsalou, 1999; Prinz, 2003). More recently, the call for an affective revolution led to a shift of attention from the rationalistic view to the empiricist view (Prinz, 2003; Greene & Haidt, 2002; Rozin et al., 1999). A result of this affective revolution was the development of the social intuitionist model (Greene & Haidt, 2002; Haidt, 2001). This model suggests

that intuitions in the form of moral emotions cause moral judgments directly because these intuitions have an affective valence, either good or bad.

Haidt (2001) described four reasons that plead for empiricism and doubt rationalism. The first reason is the dual process problem. Feelings and thoughts about a moral situation occur in parallel. But the feeling or affective evaluation of a situation occurs automatic, is pervasive, and is very quick. It seems to be an integral part of the perception of a particular (moral) situation. The intuitionist model can explain and deal with this 'problem' as opposed to the rationalistic view. The second reason is the limited working of the rationalistic model. It only works if a person has time enough, has the motivation to make an accurate decision, and is capable to make a rational and objective decision. The third reason is about the finding that people when asked why some situation is bad often give reasons that do not matter and they do not mention reasons that matter. Often they can not give a good cause for their behaviour or decision. They come up with post hoc reasons to give a plausible reason why something is good or bad. The last reason is that moral action is more related to moral emotion than to moral reasoning. Moral emotions versus moral reasoning are like the 'hot' system (amygdala-based memory) versus the 'cold' system (hippocampal memory, frontal lobe planning and inhibition areas). The 'hot' system is much quicker and the 'cold' system can block the impulses of the 'hot' system. The strength of the 'cold' system develops with age because the frontal cortex develops. All these reasons give preference to the empiricist view.

The experience of moral emotions seems necessary. In studies with psychopaths Checkley (1955, in Haidt, 2001) found they show less moral emotions. The metabolic activity in the frontal cortex was reduced in psychopaths in relation to 'normal' people (Raine, Buchsbaum, Stanley, Lottenberg, Abel, & Stoddard, 1994). Damasio, Tranel, and Damasio (1990) found that damage to the ventromedial area of the prefrontal cortex (VMPFC) lead to a pattern of affect loss. Patients reported to feel nothing when images were shown that should affect people.

It is possible to classify moral emotions into two main clusters. These clusters are related to the structure of the social environment (Rozin et al., 1999). The first cluster is directed to oneself and is sometimes called the cluster of self-conscious emotions. Emotions in this cluster are shame, embarrassment, and guilt (SEG) (Eisenberg, 2001). These emotions are crucial because they reflect the internalization of the social hierarchy. The second cluster consists of the emotions contempt, anger, and disgust (CAD). These emotions are externalizing and directed to people who do not act according to the moral rules.

Rozin et al. (1999) linked the CAD cluster of moral emotions to the "Big Three" ethics of morality (Shweder, Much, Mahapatra, & Park, 1997; Haidt, Koller, & Dias,

1993). The "Big Three" of morality is a way to class moral rules in three categories. The first category is autonomy. Rules in this category are connected to the individual and the accompanying rights and personal freedom the individual has in his or her choices. The aim of this rule is to protect the individual from harm and restriction in following someone's personal preferences. This ethic is usually the official ethic of societies where individualism is an ideal. The second category is community. Rules in this category are connected to the social hierarchy in the society. The aim is to protect the community. The community is conceived as an entity with its own identity that needs to be protected. The several duties someone has depend on the position in the society and the existing social hierarchy. This position determines the social status and the role someone has in the society. The third category is divinity. Rules in this category are related to 'the self' as spiritual entity. Rules are formed to maintain natural order, sacred order, purity, and tradition. The aim of these rules is to protect the soul, the spirit, the spiritual aspects of the individual, and to protect the natural order for degradation. The individual is seen as someone who is kept responsible with respect to (a) God. The assumption is that if a rule in a particular category is violated this will evoke a particular emotion (Rozin et al., 1999; Prinz, 2003). A specific relation between community and contempt, autonomy and anger, and divinity and disgust is supposed. This relation will not be perfect when looking at the individual level but at the general level these relations will be substantial. This supposition is called the CAD-hypothesis which refers to Community/Contempt, Autonomy/Anger, and Divinity/Disgust.

Two experiments were done to test the CAD-hypothesis. For both experiments students of a Christian College were approached. The reason to choose this population was for testing the category divinity. The expectation is that people who do not believe in (a) God will be generally more indifferent if a violation of the rules in the category divinity takes place. Of course it will depend on the kind of violation described. They will not be indifferent to all violations in the category divinity (incest or other forms of paedophilia). But if they are indifferent with respect to a violation this situation will not evoke disgust. Often they will judge such a situation not as morally bad (cursing, abortion in a later stage of the pregnancy or practising homosexuality). The opposite was expected with people who believe in God. They will not be indifferent and they will generally judge a violation of the rules in the category divinity as morally bad. According to the CAD-hypothesis they will experience disgust. If this assumption is true the chance to find a relation between divinity and disgust is the highest in this population. A Christian population and the general population do in all probability not differ from moral standard with respect to the category community and autonomy.

In Experiment 1 first an emotion was induced by a picture of a face expressing anger, disgust, or no emotion (neutral facial expression). In the first condition the picture

was shown suboptimally (16 ms) and in the second condition the picture was shown optimally (1000 ms). The presentation of the face was followed by a sentence. This sentence described a morally good or bad situation. The subject had to decide as quickly as possible whether the described situation was good or bad. The central question in Experiment 1 was: Do the three moral emotions contempt, anger, and disgust underlie a moral judgment in the category community, autonomy, and divinity respectively? According to the findings of Rozin et al. (1999) the moral judgment is expected to be faster if the category of the described situation is congruous with the induced emotion (Hermans, De Houwer, & Eelen, 1996). The kind of emotion a person is experiencing at a particular moment depends on the current mood and personality traits. Consequently the current mood and personality traits of a person will influence the reaction time of a person to make a moral judgment. To control for unwanted congruency or incongruency effects a particular emotion can be induced (Rusting, 1999).

Several studies have shown that induction of emotion by means of a picture will be more effective if an emotion is presented suboptimally in stead of optimally which is called the affective primacy hypothesis (Murphy, Monahan, & Zajonc, 1995; Murphy & Zajonc, 1993; Rotteveel, de Groot, Geutskens, & Phaf, 2001). Murphy and Zajonc (1993) investigated the affective primacy hypothesis that states: 'Affective reactions can be evoked with minimal stimulus input and virtually no cognitive processing' (p. 723). In their study neutral Chinese ideographs were preceded by either a suboptimal or an optimal affective prime. The subject had to rate how much they liked the shown ideograph. Subjects rated an ideograph as significantly higher in likeability when preceded by a positive prime compared with a negative prime in the suboptimal priming condition. In the optimal priming condition a higher rate of likeability was not found if a positive prime was used compared with a negative prime. Murphy et al. (1995) replicated the finding that affective priming was only effective in the condition with suboptimal presentation of the prime. A possible explanation is that in the suboptimal priming condition the induced emotion can not be attributed to something in the environment and will be applied to the target.

Also studies that used a more implicit measure to investigate the affective primacy hypothesis have found supporting results. Rotteveel et al. (2001) used facial electromyography (facial EMG) to investigate the affective primacy hypothesis. Facial muscles (musculus corrugator supercilii) reacted congruent to the affective prime which was presented suboptimal. Dimberg, Thunberg, and Elmehed (2000) investigated facial mimicry when the subjects were exposed to suboptimal presented happy and angry faces. Subjects responded unconsciously to both the happy and angry faces. These findings can be related to the facial-feedback hypothesis which states that facial expression is the core of the experience of an emotion (Adelmann & Zajonc, 1989). Facial

muscle activity provides feedback which is necessary or sufficient to experience an emotion (Buck, 1980). It seems no conscious cognitive process is needed to trigger an emotion (Barsalou, 2002) and for this reason suboptimal priming is an effective way to induce an emotion.

Experiment 2 was done to replicate the findings of Rozin et al. (1999) with the materials used in Experiment 1. First the subjects had to read a sentence which described a morally bad situation. In each sentence the rules of a particular moral category (community, autonomy, or disgust) were violated. Then the subjects had to choose a face that represented the emotion evoked by the sentence. The facial expressions they could choose were anger, disgust, or neutral. It was expected that subjects would choose the facial expression that was congruent to the moral category that was violated in the sentence. It follows that if the rules of the category autonomy were violated the subjects would choose an angry face, in the case of community they would choose almost random and in the case of divinity they would choose a disgusted face.

Pilotstudy

A Pilotstudy was done to select materials for the experiments that follow. The materials needed are sentences that describe a morally bad situation in a particular category (community, autonomy, divinity) and faces to serve as prime to induce a particular emotion (anger, disgust, neutral). Sentences that were most consistently grouped in a particular category were selected for the experiments. The faces were selected from the Karolinska set (Lundqvist, Flykt, & Öhman, 1998) according to the most consistent rated emotion and the rated strength of the represented emotion.

Method

Participants. The subjects were 16 undergraduate psychology students at the Erasmus University in Rotterdam who participated for course credit.

Materials. Two sets of materials were created. The first set consisted of 120 sentences all describing a morally 'bad' situation. In each sentence the rules of a particular category (autonomy, community, or divinity) were violated. The sentences were developed according to the description of the categories by Shweder et al. (1997). The keywords used as guideline for each category are displayed in Table 1. The sentences used to represent the category divinity differed in one respect to Rozin et al. (1999). They used sentences as: 'a person is eating a piece of rotten meat' or 'a person is touching a corpse' (p. 578). Regardless from the moral judgment of the described situations these situation would evoke disgust. Not only because they fit in the category divinity but also (and perhaps more) because they are physically disgusting.

Table 1

Guidelines used to develop the sentences to represent each category

Autonomy	Community	Divinity
Harm, Rights, Justice	Duty, Hierarchy, Interdependence, Souls	Sacred Order, Natural Order, Sanctity, Tradition
Individual as a preference structure	Actor in a play Role-based social status Family	Way of life Practice World-Soul
Obligations come from being a person	Obligations come from being part of a community	Displaying dignity by showing ultimate concerns
Free agent	Social, not just selfish	Human, not beast

Source: Shweder et al. (1997)

To control for this the developed sentences satisfied the description of the category divinity but were not physically disgusting. Each category was represented by 40 sentences. The participants had to categorize these 120 sentences in one of the three moral categories. The second part of the Pilotstudy consisted of 120 faces (angry, disgusted, or neutral). These faces all came from the Karolinska set and were thus validated (Lundqvist et al., 1998).

Procedure. The Pilotstudy was programmed in E-Prime and performed on the computer. First the participants had to read a brief description of both parts of the experiment. Then specific instructions for the first part appeared. They had to read a description of the categories (autonomy, community, and divinity) each representing a set of rules that could be violated. They had the task to decide which kind of violation was described by the sentence. The sentences were displayed on a 17 inch monitor in front of the participant. Below the sentences the response categories were displayed. The response categories were: autonomy, community, divinity, other, or they could choose to have a look at the description of the categories again before they gave their response. They had to press the 1, 2, 3, 4, or 9 on the keyboard respectively. The participants had the possibility to read the description of the categories whenever they needed or wanted. After reading the description they could return to the sentence. After finishing the first part they had to read specific instructions for the second part. In the second part they saw a face and an open-ended question was used so as not to force the participants to choose one of the three emotions (anger, disgust, or neutral). They could type their answer in a specified area below the face. After pressing 'Enter' the participants saw the same face for the second time. They were asked to rate the strength of the displayed emotion on a three item scale: not so strong, a bit strong, and strong. They had to press the keys 1, 2, or 3 respectively. Both parts of the experiment were not time limited. The sentences and faces in the multiple choice conditions were shown until the participants had chosen one of the valid response categories.

Experiment 1

After conducting the Pilotstudy the materials (sentences and faces) needed for the experiments were collected. Experiment 1 was designed to test the CAD-hypothesis by means of suboptimal and optimal affective priming. The expectation was that the mean reaction time to judge whether the described situation is good or bad would be shorter if the emotion induced by the picture of the face was congruent to the emotion evoked by the sentence.

Method

Participants. In this study 84 second year students from a Christian College (The Driestar in Gouda) participated. The most distinctive feature of this school is that the staff members and the students adhere to the Reformed faith. All the participants received training to become teachers on a (Christian) primary school. They received € 4,- for participating. They had the choice to receive it cash or to donate it to a charity.

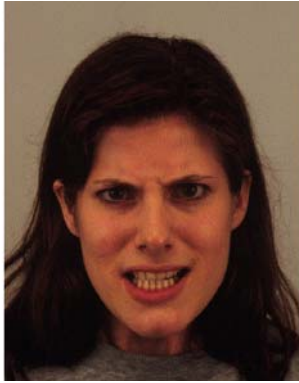
Materials. Based on the first part of the Pilotstudy 90 sentences were selected. Each category (autonomy, community, and divinity) was represented by 30 morally 'bad' sentences. As fillers were used 90 morally 'good' sentences. The fillers were matched as much as possible on subject and length with the morally 'bad' sentences.

Based on the second part of the Pilotstudy 20 faces of each kind of emotion (anger, disgust, and neutral) were selected. With a total of 180 sentences and 60 faces each face was shown three times during the experiment. The faces of 32 women and 28 men were used. From each of the selected faces a very blurred form was made in Microsoft Picture It! to serve as a mask. First the face was blurred for 300% and second an illusion was used to make the represented emotion unrecognizable. Some examples of the faces and accompanying masks are displayed in Figure 1.

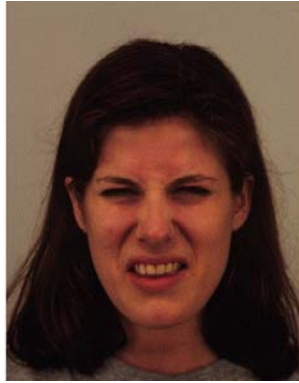
The pictures, masks, and sentences were displayed on a flatscreen in front of the participants. The display settings were 1024 by 768 pixels, a screen refresh rate of 60 Hz, and the use of the highest possible colour quality (32 bits). The distance between the subject's eyes and the flatscreen was about 45 cm. The size of both the pictures and the masks was 11.8 cm (width) by 15.1 cm (height).

Procedure. The experiment consisted of two conditions: a suboptimal and an optimal condition. Each condition consisted of three lists. These lists were designed to fully counterbalance the experiment. Each sentence was preceded by each kind of emotion (anger, disgust, neutral). The participants were randomly assigned to one of the two conditions and then randomly assigned to one of the three lists.

First the participants saw a display with instructions. They had to place their right indexfinger on the 'm'-key and their left indexfinger on the 'x'-key. The 'x'-key represented morally 'bad' and the 'm'- key morally 'good'.



a) Anger



b) Disgust



c) Neutral



d) Mask anger



e) Mask disgust



f) Mask neutral

Figure 1. Examples of the faces and accompanying masks used.

After receiving those instructions nine exercise trials were followed by 180 experimental trials. A trial consisted of the picture, which was shown for 16 ms in the suboptimal condition and for 1000 ms in the optimal condition. The choice of 16 ms was because of the screen refresh rate. The duration of one refresh cycle is at least $1000/60 \text{ Hz} = 16,67 \text{ ms}$. Shorter presentation times were not possible. After the picture the mask appeared for 1000 ms in both conditions. After showing the picture the sentence was shown without a time limit. The participants had to decide as quickly as possible whether the described situation was morally 'good' or 'bad'. The sentences were presented randomly.

Results

First an outlier analysis was done to remove outlier reaction times. All reaction times that were more than two standard deviations above or below the mean reaction time of a person were removed. This procedure was repeated for every participant. The total number of removed reaction times corresponded to 4.66% of all the reaction times ($352/7560 \approx 0.0466$).

Analysis of reaction time. A 3 x 3 ANOVA was performed to compare the subject's reaction time across the within-subjects factor category (autonomy, community, and divinity) and the within-subjects factor emotion (anger, disgust, and neutral). The results revealed no significant category x emotion interaction effect in both the suboptimal condition, $F(4, 38) = .483, p = .748$ and the optimal condition, $F(4, 38) = .736, p = .573$. The mean reaction time was not significantly shorter if the primed emotion was congruent to the emotion evoked by the described situation. The kind of affective prime (anger, disgust, or neutral) seemed not significantly related to the kind of category (autonomy, community, or divinity). This finding did not support the CAD-hypothesis. The results revealed also no significant main effect for emotion in both the suboptimal condition, $F(2, 40) = .982, p = .384$ and the optimal condition, $F(2, 40) = .288, p = .751$. No significant differences in reaction times were found related to the specific emotions represented by the pictures. There was a significant main effect for category in the suboptimal condition, $F(2, 40) = 5.784, p < .001$. As represented in Figure 2 the mean reaction time for sentences in the category divinity was substantially shorter than the mean reaction times of the other two categories regardless of the primed emotion. No significant main effect for category was found in the optimal condition, $F(2, 40) = .865, p = .429$. The mean reaction times in the optimal condition are represented in Figure 3.

Erroranalysis. Not only the reaction time but also the amount of errors could be used to test the CAD-hypothesis. Accuracy was expected to be higher in the congruent condition than in the incongruent condition. To examine this prediction a 3 x 3 ANOVA was performed. The accuracy of the subjects was compared across the within-subjects factor category (autonomy, community, and disgust) and the within-subjects factor emotion (anger, disgust, and neutral).

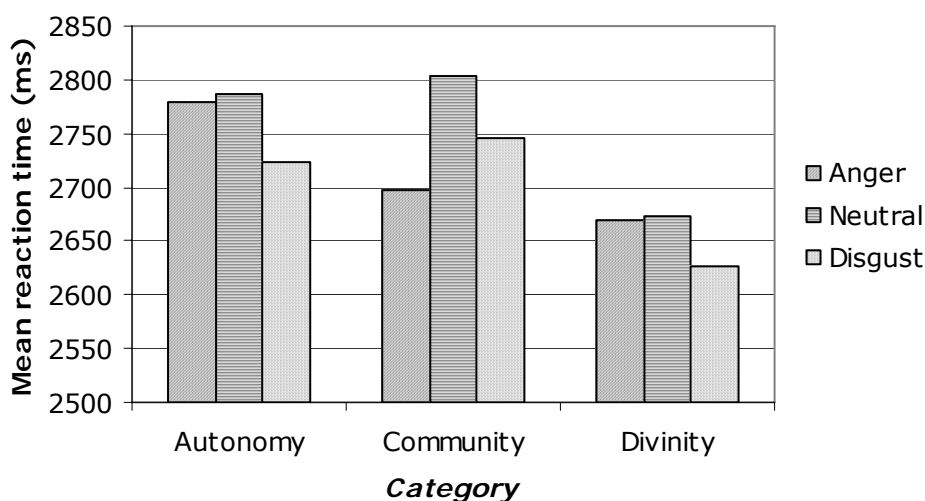


Figure 2. Mean reaction times in the suboptimal condition.

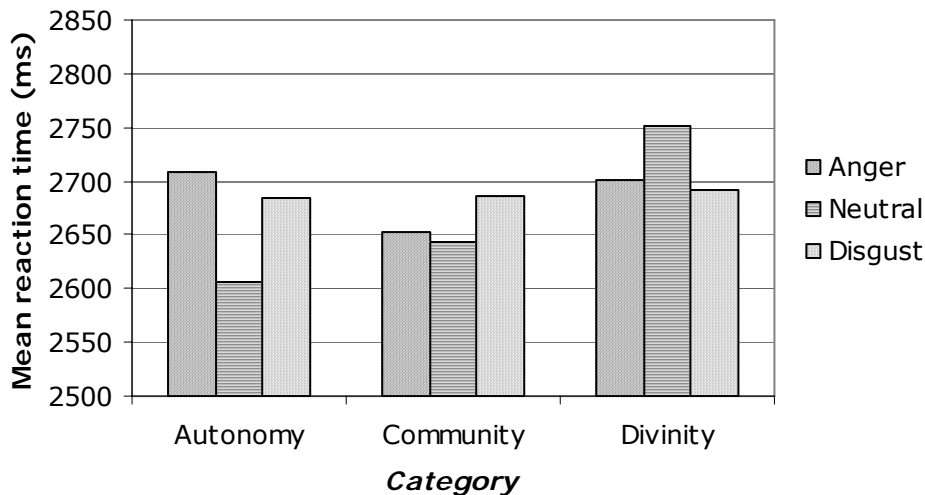


Figure 3. Mean reaction times in the optimal condition.

The category x emotion interaction effect was not significant in the suboptimal condition, $F(4, 38) = 1.427, p = .244$ and in the optimal condition, $F(4, 38) = .298, p = .877$. The kind of emotion induced by the prime (either suboptimally or optimally) was not significantly related to the accuracy of the judgment. In other words the accuracy was not significantly higher if the emotion represented by the picture resembled the emotion evoked by the sentence as should be expected according to the CAD-hypothesis. A significant main effect was found for category in the suboptimal condition, $F(2, 40) = 12.358, p < .001$ and the optimal condition, $F(2, 40) = 7.565, p < .005$ but not for emotion in the suboptimal condition, $F(2, 40) = .542, p = .586$ and the optimal condition, $F(2, 40) = .462, p = .633$. The mean accuracy of the category divinity was higher regardless of the affective prime. The mean accuracy for each combination of category and emotion in the suboptimal condition and the optimal condition is shown in Figure 4 and Figure 5 respectively.

Additional analyses. The two analyses described above both did not support the CAD-hypothesis. Before drawing conclusions other variables that might have influenced the results had to be taken into account. Perhaps the faces did not induce the emotion they were meant to induce and for this reason no shorter mean reaction times were found in the congruent situations. This seems not plausible because all the faces came from a validated set of pictures (Lundqvist et al., 1998). In addition a Pilotstudy was done to select the 'best' faces from this set. Another subject of discussion could be the sentences.

The first possibility was that sentences with a low accuracy distorted the results because participants rated a morally 'bad' situation as a 'good' one. To examine this all the sentences which had an accuracy lower than .90 were left out of the analysis.

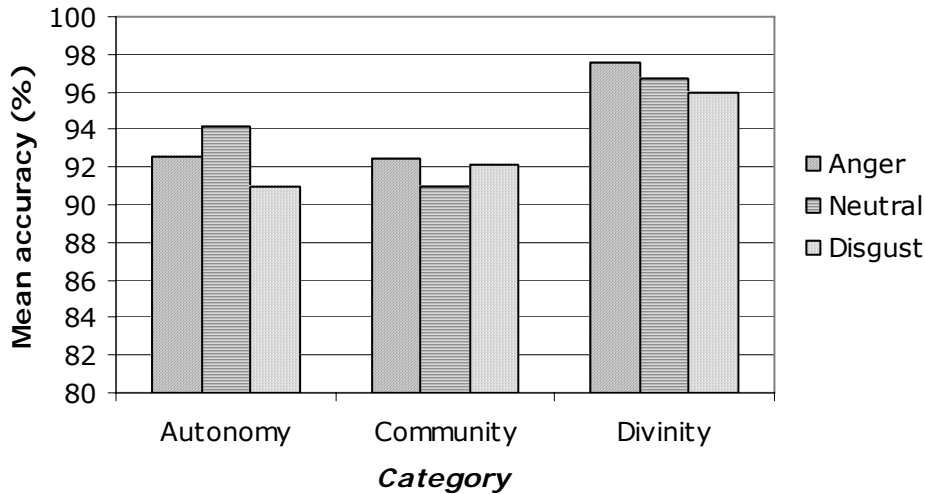


Figure 4. Mean accuracy in the suboptimal condition.

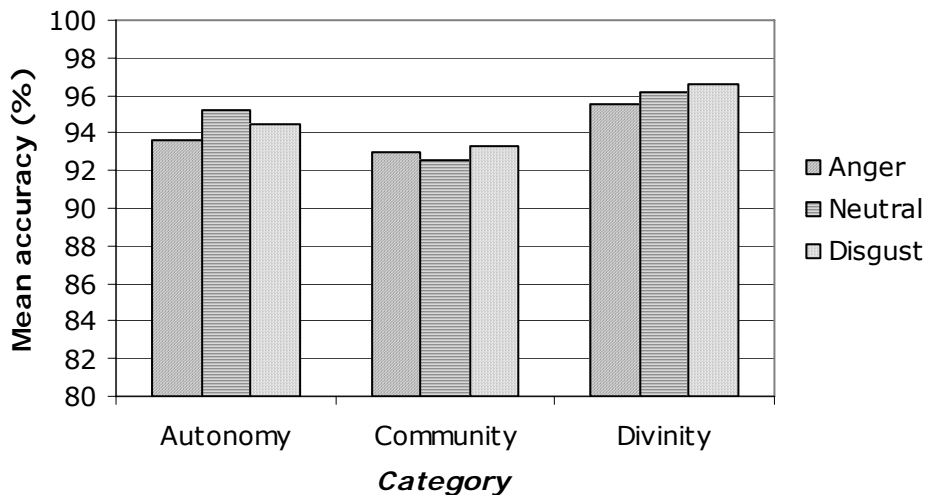


Figure 5. Mean accuracy in the optimal condition.

In sum 18 sentences out of 180 were omitted. The reaction times were analyzed with a 3 (category x 3 (emotion) ANOVA. This interaction was not significant in both the suboptimal condition, $F(4, 38) = .788, p = .540$ as in the optimal condition, $F(4, 38) = .995, p = .422$. No significant main effect for category was found in both the suboptimal condition, $F(2, 40) = 1.983, p = .151$ and optimal condition, $F(2, 40) = 1.192, p = .314$. Also no significant main effect was found for emotion in both the suboptimal condition, $F(2, 40) = .499, p = .611$ and the optimal condition, $F(2, 40) = .609, p = .549$. Thus, removal of sentences with low accuracy did not change the pattern of results.

The second possibility might be the accuracy of the category of the sentences. For example a sentence was meant to represent a situation in the category autonomy but in

the eyes of the participants it represented a situation in the category community. This could happen because the population participating in the Pilotstudy were not the same as the population participating in Experiment 1. To control for this possibility 10 students of the Driestar (same study, same year) performed the Pilotstudy. All the sentences that were for 50% or less categorized in the right category were removed. A total of 22 sentences were removed of which 10 in the category autonomy, 10 in the category community and 2 in the category divinity. A 3 (category) x 3 (emotion) ANOVA revealed no significant interaction effect, $F(4, 38) = .350, p = .843$ and no significant main effect for category, $F(2, 40) = 1.709, p = .19$ and for emotion, $F(2, 40) = 1.240, p = .300$ in the suboptimal condition. The results of the optimal condition revealed also no significant interaction effect, $F(4, 38) = .869, p = .491$ and no significant main effect for emotion, $F(2, 40) = .215, p = .807$. But for category a significant main effect was found, $F(2, 40) = 5.663, p < .001$. Several times a significant main effect for category was found but the results were mixed. In the suboptimal condition the reaction times of judgments in the category divinity were substantially shorter and in the optimal condition the reaction times were longer. Perhaps someone will put forward that the length of the sentences could have influenced the results. But this was not likely because the design was fully counterbalanced. Also additional analyses did not support the predictions made by the CAD-hypothesis.

Experiment 2

In Experiment 1 no evidence was found for the CAD-hypothesis as opposed to the results of Rozin et al. (1999). Experiment 2 was designed to replicate study 1 of Rozin et al. (1999) and to examine whether their results would be found with the materials used in Experiment 1. The participants had to decide which face (emotion) most likely will be shown by a person who observed the described morally 'bad' situation.

Method

Subjects. In this study 76 students of The Driestar in Gouda participated. They were all first and second year students and from the same study as in Experiment 1. They received € 4,- for participating. They had the choice to receive it cash or to donate it to a charity.

Materials. The same 90 sentences that described a morally 'bad' situation and the same 60 faces (20 angry, 20 disgusted, and 20 neutral faces) as in Experiment 1 were used. The sentences and faces were shown on a flatscreen (17 inch) in front of the participant. The distance between the participant and the display was about 45 cm. The display settings were 1024 by 768 pixels, a screen refresh rate of 60 Hz, and the highest possible colour quality. The size of the pictures was 10.8 cm (width) by 13.5 cm (height).

Procedure. The experiment was programmed in E-prime and performed on the computer. The first two displays were meant to describe the instructions of the experiment. The instructions were similar to the instructions of study 1 of Rozin et al. (1999). After pressing 'Enter' the experiment started. A sentence was shown in a frame. Beneath the sentence three faces were shown with under each face a number (1, 2 or 3). The answer possibilities were 1, 2, or 3. The trials were not time limited. If the participants gave an answer the next trial appeared with a total of 90 trials.

Three different lists were created. The participants were randomly assigned to one of the three lists. The lists were designed to counterbalance the experiment. Every sentence was shown in combination with a different order of faces to control for order effects. For example, a sentence in the first list was shown with from the left to the right an angry, a neutral, and a disgusted face. In list two the same sentence was shown with a neutral, a disgusted, and an angry face respectively and in the third list this sentence was shown with a disgusted, an angry, and a neutral face respectively. The sentences were shown randomly.

Results

Analysis of choice. The aim was to compare the subject's choice (anger, neutral, and disgust) across the within subjects factor category (autonomy, community, and divinity). First for each possible combination of category and emotion the mean number this combination was chosen was calculated. Then a 3 (category) x 3 (emotion) ANOVA was performed. A significant interaction effect was found, $F(4, 72) = 43.215, p < .001$. Also a significant main effect for emotion was found, $F(2, 74) = 38.885, p < .001$. The main effect for emotion could be explained by the high frequency the angry face was chosen. It seemed people experience most often the emotion anger when they see another person violating a moral rule. The main effect for category was not significant, $F(1, 75) = 1.000, p = .321$. As shown in Figure 6 the angry face was chosen the most if the situation described in the sentence violated a rule in the category autonomy. Two paired samples t-tests were done to investigate whether anger was chosen significantly more than disgust or neutral if a rule in the category autonomy was violated. The results were significant at a .001 level. The same was done for disgust and divinity. But the paired samples t-tests revealed no significant results. The disgusted face was not chosen significantly more if a rule in the category divinity was violated. The emotion chosen was almost random because the proportion of choices for each emotion was around 33%. As opposed to the expectation if a rule in the category community was violated the emotion chosen was not random. Participants chose in almost half of these situations the angry face.

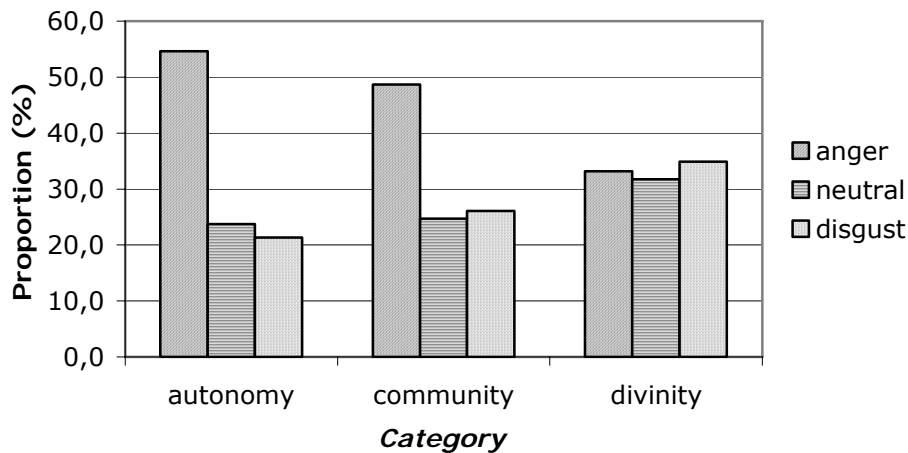


Figure 6. Proportion of choices of an emotion in a particular category.

Analysis of reaction time. Second an analysis of reaction times was done. The participants were not forced to choose the best fitting face as fast as possible. But nevertheless the reaction time for every choice was recorded. A 3 (category) x 3 (emotion) ANOVA was performed to compare the mean reaction times. This analysis revealed no significant interaction effect, $F(4, 72) = .567, p = .687$. The main effect for category was also not significant, $F(2, 74) = 2.863, p = .063$. But there was a significant main effect for emotion, $F(2, 74) = 10.302, p < .001$. Figure 7 displays the mean reaction times. Notable were the 'long' reaction times before the subjects chose the neutral face which explained the main effect for emotion. It seemed to be difficult for the participants to choose the best fitting emotion. Also notable were the mean reaction times if the sentence described a violation in the category community. The mean reaction times were longer regardless of the emotion the participants chose.

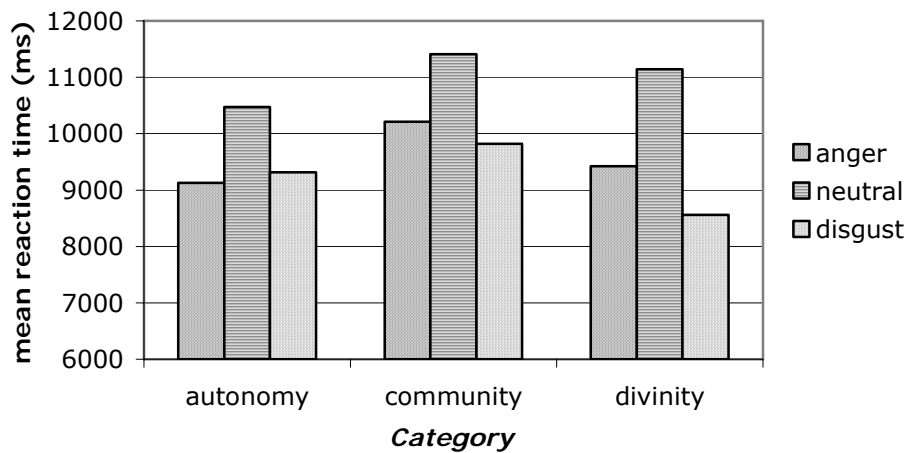


Figure 7. Mean reaction times to choose an emotion in a particular category.

According to the results of the ANOVA this difference was not significant because no significant main effect for category was found. If a rule in the category autonomy was violated, the mean reaction time before the participants chose the angry face was the shortest. To look if the mean reaction time was significantly shorter two paired samples t-tests were done. Both the difference between autonomy-anger and autonomy-neutral and the difference between autonomy-anger and autonomy-disgust were not significant. If a rule in the category divinity was violated the mean reaction time to choose a face expressing disgust was the shortest. Again two paired samples t-tests were done. The difference between divinity-disgust and divinity-neutral was significant ($p < .001$) but the difference between divinity-disgust and divinity-anger was not significant.

Additional analysis. For each sentence the mean number a particular emotion was rated was analyzed. In the category autonomy and divinity 15 sentences with the highest rated congruent emotion were selected. The relation between category and emotion was most evident in these sentences according to the CAD-hypothesis. The mean reaction times of these sentences in Experiment 1 were analyzed again. First an outlier analysis was done to remove reaction times that were more than two standard deviations above or below the mean reaction time of a participant. Of the total number of reaction times 4.72% was removed (4.76% in the suboptimal condition and 4.68% in the optimal condition). A 2 x 3 ANOVA was performed to compare the within-subjects factor category (autonomy and divinity) with the within-subjects factor emotion (anger, neutral, and disgust). No significant interaction effect was found in the suboptimal priming condition, $F(2, 40) = .074, p = .929$. This analysis revealed also no significant main effect for category, $F(1, 41) = 5.406, p = .025$ and for emotion, $F(2, 40) = .233, p = .794$. In the optimal priming condition the interaction effect for category x emotion was also not significant, $F(2, 40) = 2.222, p = .122$. But the main effect for category was significant, $F(1, 41) = 14.407, p < .001$.

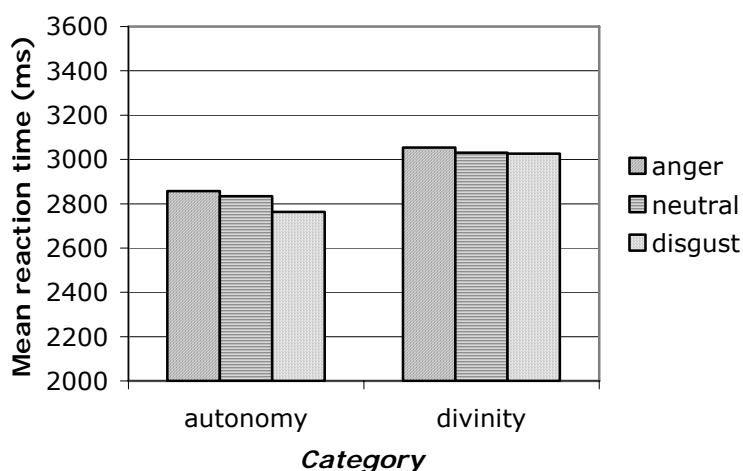


Figure 8. Mean reaction times in the suboptimal condition.

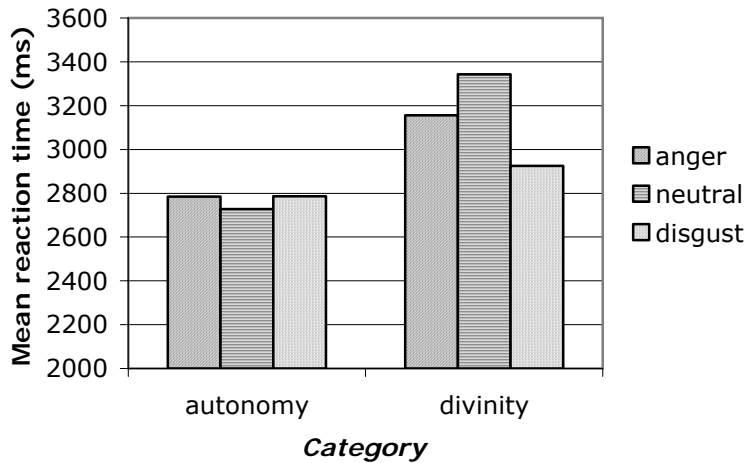


Figure 9. Mean reaction times in the optimal condition.

The mean reaction times were longer when the sentence described a violation in the category divinity regardless of the induced emotion. No significant main effect for emotion was found, $F(2, 40) = 1.277, p = .290$. The results for the suboptimal condition are displayed in Figure 8 and the results for the optimal condition in Figure 9. It should be pointed out that both the results of the suboptimal and the optimal condition did not support the CAD-hypothesis. Even if the sentences with the highest evidence for congruency between category and emotion were selected, evidence for the CAD-hypothesis was not found.

Discussion

The data of Experiment 1 did not support the CAD-hypothesis. No significant shorter mean reaction times were found if the emotion induced (suboptimally or optimally) was congruent to the emotion evoked by the described situation. No substantial relation is found between autonomy and anger and divinity and disgust in Experiment 1. Experiment 2 only partly supported the CAD-hypothesis. A significant relation was found between autonomy and anger but not between divinity and disgust. Looking at the sentences which described a violation of the rules in the category community the angry face was also chosen the most. Thus the results seemed not specific for the relation between autonomy and anger. In other words the data of both Experiment 1 and Experiment 2 did not support a specific relation between a violation in a particular category and the experience of a particular emotion.

Experiment 1 was done to investigate whether a particular emotion underlies the judgment of a particular moral situation. Prinz (2003) supposed morality is grounded by experiencing a particular emotion. The emotion someone is experiencing is a reflection of a change in the perceptual state of the body. This perceptual state was induced by

showing a facial expression. But no specific relation was found between the experience of anger and disgust and a violation in the category autonomy and divinity respectively. Because no face which expresses contempt was used no conclusions can be drawn about the category community. The expectation that the effect of congruency of category and emotion would be stronger in the suboptimal priming condition than in the optimal priming condition was also not supported. Both the suboptimal and the optimal condition did not reveal significant interaction effects between category and emotion. If the CAD-hypothesis would be substantial looking at the group level the expectation was that the chance to find evidence was the highest in a Christian group. Because they believe in (a) God and they would experience an emotion if a violation against God was described. Several sentences in the category divinity represented a violation against God. But even in this group the specific relation between divinity and disgust was not supported and for this reason no evidence for the CAD-hypothesis was found.

The structure of Experiment 1 was different from the experiment (study 1) done by Rozin et al. (1999). Perhaps this might be the reason no evidence was found for the CAD-hypothesis as opposed to the results of Rozin et al. (1999). To rule out this possibility Experiment 2 was done to replicate study 1 of Rozin et al. (1999). But even under these circumstances the CAD-hypothesis was only partly supported. If a sentence described a violation in the category autonomy the participants chose the face expressing anger significantly more than the other two emotions. The same pattern was observed if a rule in the category community was violated. If a rule in the category divinity was violated the emotion chosen was almost random. This was not in accordance with the expectation based on the CAD-hypothesis. The 15 sentences with the highest congruency between category and emotion in the category autonomy and in the category divinity were selected. The results of these sentences in Experiment 1 were analyzed again. But no support for a specific relation between autonomy and anger and divinity and disgust was found in both the suboptimal and the optimal condition.

Before coming up with conclusions it is important to have a look at how these results are obtained. Perhaps the designs of the experiments were not the right ones to measure the CAD-hypothesis. This seems not probable. First, both the designs of Experiment 1 and 2 were fully counterbalanced. Second, because additional analyses were done to rule out several possibilities that could account for the obtained results. Additional analyses done in Experiment 1 were removing sentences that were for 10% or more judged as good while they were meant to be bad and sentences that were for 50% or less categorized in the category they should represent. In Experiment 2 the 15 sentences with the highest rated congruent emotion in the category autonomy and divinity were selected and the mean reaction times of these sentences in Experiment 1 were analyzed again. But none of these additional analyses revealed significant results

that supported the CAD-hypothesis. Third, the structure of the experiments might have played a role. In Experiment 1 first an emotion was induced by means of a picture. Induction of emotion by showing a facial expression has been proven effective in several studies (Murphy et al., 1995; Murphy & Zajonc, 1993; Rotteveel et al., 2001; Wild, Erb, & Bartels, 2001). Particularly suboptimal priming of emotion revealed significant inducement of emotions. The priming stimulus was followed by a mask and after the mask a sentence appeared. If moral judgments are grounded by experiencing a particular emotion or emotional valence these two emotions (evoked by the picture and evoked by the sentence) can be congruent or incongruent. Based on the rules of congruency the reaction time will be shorter if the induced emotion is congruent to the emotion the sentence evokes (Hermans et al., 1996). It follows that the design of Experiment 1 was an effective one to test the CAD-hypothesis and to measure the grounding of morality.

The design of Experiment 2 was the same as study 1 done by Rozin et al. (1999). The only differences were the materials. The 'materials' might be the second point of discussion of how the results are obtained. Perhaps the materials used in these experiments may account for the results. That sounds not very convincing. First, the faces used in these experiments came all from a validated set of pictures (Lundqvist et al., 1998). In addition a Pilotstudy was done to select the 'best' faces. Rozin et al. (1999) used faces that were judged by a certified Facial Action Coding System rater. Second, the sentences were developed according to the description of the categories by Shweder et al. (1997). A Pilotstudy was done to select the sentences which were most consistently grouped in the category they were assumed to represent. The only difference between the experiment of Rozin et al. (1999) and the materials used in these experiments was the absence of physically disgusting sentences. This seems only a strong point of these experiments. Because if a sentence represents a violation of a rule in the category divinity this violation has to evoke disgust, regardless of the described situation is physically disgusting or not according to the CAD-hypothesis. The masks resembled the primes very much. The diffusion of colours was the same, only the represented emotion was made unrecognizable. The possibility of using the 'wrong' materials can be ruled out.

The third point which can be discussed with regard to how the results are obtained is the population which participated. This population is not representative for the general population. It is a Christian group with strict norms but their moral standards or degree of morality will not be higher than the general population. It seems plausible that for the categories autonomy and community there will be no substantial differences between a Christian population and the general population. The relation between divinity and disgust was expected to be more substantial in this population than in the general

population. Although using a Christian population the relation between disgust and divinity was not found.

Is it possible considering the results of these two experiments to draw the conclusion that the CAD-hypothesis does not make a right prediction? The answer is a straightforward 'no'. There are several reasons that must be taken into account. The first reason is that in this study no contemptuous face was used. Afterwards it is very difficult to predict the influence of using a contemptuous face instead of a neutral face. If in Experiment 2 a contemptuous face was used instead of a neutral face the participants were forced to choose an emotion after reading the sentence. When it was difficult for the participants to choose the best fitting emotion they had the opportunity to choose the neutral face. The mean reaction time before the participants chose the neutral face was significantly longer than the mean reaction time before the participants chose the angry or the disgusted face. Perhaps if the neutral face was not an option someone can speculate that the relation between community-contempt and divinity-disgust should become clearer. In agreement with this were the slightly longer mean reaction times before a face was chosen after reading a sentence which described a violation in the category community. Taken all these considerations into account they can not explain why no specific relation between divinity and disgust was found. Based on the CAD-hypothesis for all the three categories a substantial relation with an emotion was expected to be found.

The second reason for not rejecting the CAD-hypothesis is the difference in the sentences used to represent the category divinity. The sentences used in this study and the sentences used by Rozin et al. (1999) both satisfied the description of the categories by Shweder et al. (1997). But they differed from a very influencing point. A part of the sentences used by Rozin et al. (1999) were physically disgusting. Seeing someone eating an apple with a worm in it will evoke disgust in all of us regardless of the moral value. It is not in the first place a matter of doing something which is morally bad. In the study of Rozin et al. (1999) the category divinity is represented by five sentences of which three sentences describing a physically disgusting situation. The mean number the disgusted face was chosen was much higher if the described situation was physically disgusting (92%, 88%, and 74%) than less or not physically disgusting (40% and 59%). In this study thirty sentences were used which described no physically disgusting situations. Because they satisfied the description by Shweder et al. (1997) they should evoke disgust according to the CAD-hypothesis. But the proportions found were almost random. Each face was chosen around 33%. Perhaps the physically disgusting situations were the reason Rozin et al. (1999) found a relation between divinity and disgust. If this is true the results of Rozin et al. (1999) can not be taken as absolute evidence for a relation between divinity and disgust.

The third reason no support for the CAD-hypothesis was found could be attributed to the structure of Experiment 1. The assumption made in Experiment 1 is that the experience of an emotion underlies the moral judgment. Rozin et al. (1999) talk about 'a linkage between' or 'will often be triggered by' (p. 576). They do not suppose the experience of anger, contempt, or disgust underlies a moral judgment if a rule in the category autonomy, community, or divinity respectively is violated. Perhaps the linkage between the emotion and the category will be substantial at a later stage of the moral reasoning process. This reason can not explain the results of Experiment 2 in which the best fitting emotion had to be chosen. It is possible to go one step further and propose making up a moral judgment does not depend on the experience of a particular emotion. By proposing this the grounding of morality is called into question.

Does the results of these experiments point out that emotions do not underlie moral decisions? And that moral judgments are not grounded by experiencing an emotion? Again a straightforward 'no' can be the answer. The CAD-hypothesis makes very specific predictions. According to the social intuitionist model the emotion someone is experiencing, has a valence, either 'good' or 'bad' (Green & Haidt, 2002; Haidt, 2001). It is possible no specific emotion underlies the moral judgment but a more general feeling of approval or disapproval (Hume, 1751/1983; Prinz, 2003). The perceptual symbol theory does not make specific predictions with regard to the grounding of morality (Barsalou, 1999; Barsalou et al., 2002). A better way to investigate the perceptual symbol theory would be by inducing positive and negative emotions. It is possible incongruency does not take place if the emotions anger, disgust, or neutral will be used to induce an emotion and will be compared with each other. Replicating Experiment 1 using faces with a positive valence (happy face) and a negative valence (angry face) is highly recommended. The aim would not be to compare the different categories and the related emotions but to compare morally 'good' and 'bad' situations after induction of either a face which expresses a positive or a negative emotion. If moral judgments are based on the experience of an emotion the mean reaction time will be shorter if the valence of the described situation is congruent to the induced emotion (Hermans et al., 1996).

Based on the results of Experiment 1 it is possible to conclude that the emotions anger and disgust did not underlie a moral judgment in the category autonomy and divinity respectively. According to the results of Experiment 2 it seems a violation in the category autonomy evoked anger in a later stage of the decision process, but the same was true for a violation in the category community. Further research is necessary to get more clarity. Future directions are to replicate Experiment 1 and 2 by using a contemptuous face instead of a neutral face and replicating Experiment 1 by using faces which either a positive or negative emotional valence.

Acknowledgments

The author is grateful to Diane Pecher for her valuable input and support. I also want to thank the Driestar in Gouda for giving permission to administer these experiments on their school and the students for participating in this study.

References

- Adelmann, P. K. & Zajonc, R. B. (1989). Facial efference and the experience of emotion. *Annual Review of Psychology, 40*, 249-280.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.
- Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. A. (2002). Social embodiment. In B. H. Ross (Ed.), *The psychology of learning and motivation*. San Diego, CA: Academic Press. Under review.
- Buck, R. (1980). Nonverbal behavior and the theory of emotion: The facial feedback hypothesis. *Journal of Personality and Social Psychology, 38*, 811-824.
- Damasio, A., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioral Brain Research, 41*, 81-94.
- Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11*, 86-89.
- Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology, 51*, 665-697.
- Gibbs, J. C. (2003). *Moral development and reality: Beyond the theories of Kohlberg and Hoffman*. Thousand Oaks: Sage Publications.
- Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*, 517-523.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834.
- Haidt, J., Koller, S. H., & Dias M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*, 613-628.
- Hermans, D., De Houwer, J., & Eelen, P. (1996). Evaluative decision latencies mediated by induced affective states. *Behaviour Research and Therapy, 34*, 483-488.
- Hume, D. (1751/1983). *An enquiry concerning the principles of morals*. Indiana: Hackett Publishing Company.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of Socialization Theory and Research* (pp. 347-480). Chicago: Rand McNally & Company.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. New York: Basic Books.
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces.

- Murphy, S. T., Monahan, J. L., & Zajonc, R. B. (1995). Additivity of nonconscious affect: Combined effects of priming and exposure. *Journal of Personality and Social Psychology, 69*, 589-602.
- Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology, 64*, 723-739.
- Piaget, J. (1932/1965). *The moral judgement of the child*. (M. Gabain, Trans.) New York: Free Press.
- Prinz, J. J. (2003). Passionate thoughts: The emotional embodiment of moral concepts. In R. Zwaan & D. Pecher (Eds.), *The grounding of cognition: The role of perception and action in memory, language, and thinking* (pp. 93-114). Cambridge: Cambridge University Press.
- Raine, A., Buchsbaum, M. S., Stanley, J., Lottenberg, S., Abel L., & Stoddard, J. (1994). Selective reductions in pre-frontal glucose metabolism in murderers. *Biological Psychiatry, 36*, 365-373.
- Rotteveel, M., De Groot, P., Geutskens, A., & Phaf, R. H. (2001). Stronger suboptimal than optimal affective priming? *Emotion, 1*, 348-364.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76*, 574-586.
- Rusting, C. L. (1999). Interactive effects of personality and mood on emotion-congruent memory and judgment. *Journal of Personality and Social Psychology, 77*, 1073-1086.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "Big Three" of morality (autonomy, community, divinity) and the "Big Three" explanations of suffering. In A. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119-169). New York: Routledge.
- Sullivan, R. J. (1989). *Immanuel Kant's moral theory*. Cambridge: Cambridge University Press.
- Wild, B., Erb, M., & Bartels, M. (2001). Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences. *Psychiatry Research, 102*, 109-124.