

# SCIENTIFIC REPORTS



OPEN

## Individual differences in the early recognition of moral information in lexical processing: An event-related potential study

Qun Yang<sup>1,2</sup>, Canhuang Luo<sup>1,2</sup> & Ye Zhang<sup>1,2</sup>

Previous studies have shown that intuitive moral cognition occurs at an early stage. However, inconsistent findings indicate that moral information is recognized at a relatively late stage. This study uses the recognition potential (RP) as a neural index and simultaneously measures individuals' moral preferences using the Moral Foundation Questionnaire. We aim to investigate how individual differences in moral preferences modulate the processing of morality in the pre-semantic stage and provide some insights to explain the variation in rapid information processing linked to morality. The participants performed an implicit task in which recognizable words depicting geographical names or behaviors related to moral, disgusting or neutral content alternated with background stimuli at high rates of presentation. The results showed that the early recognition of moral information manifested in the RP depended on an individual's moral concerns. Participants with a higher level of endorsement of the harm/care foundation exhibited a greater net moral effect, namely, greater mean amplitudes of the moral-neutral RP difference waves. Meanwhile, only the group that was more sensitive to the harm/care foundation showed a distinctively larger RP for the moral words than for the neutral words. Overall, these findings suggest that the early processing of moral cognition may hinge on individual differences in moral concerns about other people's suffering.

In traditional moral psychology, moral judgment is featured as a process that involves conscious mental activities and verbal reasoning<sup>1</sup>. Despite rich evidence for the influence of reasoning on moral judgment, a greater number of studies have highlighted the role of intuitive processes<sup>2</sup>.

Having the advantage of high temporal resolution, the event-related potential (ERP) technique has the ability to provide crucial insights into the online processing of moral intuition. Recently, an increasing number of ERP studies have emerged, revealing that different ERP components are associated with early stages of moral cognition. An initial study reported that a larger positivity peaking at approximately 200 ms was observed when participants made rescuing decisions between two relatives than between two strangers endangered by an earthquake<sup>3</sup>. Similarly, a larger P200 was induced by the more unpleasant moral scenarios in another study when the subjects made decisions on two types of moral dilemmas. In addition to the findings that valence ratings on unpleasantness during decision-making are significantly correlated with the P200 amplitudes, the authors noted that this early component may reflect affective processing during the first phase of moral decision-making<sup>4</sup>. A very recent study used more probable moral transgression vignettes and compared the temporal processing of social-moral violations to that of knowledge-based violations. A relatively late positivity was observed starting at 320 ms after the stimulus onset uniquely associated with spontaneous moral evaluation<sup>5</sup>.

In addition, a distinction between prosocial and antisocial behaviors is associated with early components, such as N1 and N2. In particular, behavioral ratings of moral blame have been reported to be negatively correlated with N2 amplitudes<sup>6</sup>. This early moral effect has also been reported in preschool children—early posterior negativity (EPN) was evoked while children watched cartoon characters engaging in moral behaviors, with greater negativities for helping than for harming scenes, and automatic attentional or emotional response was considered to be involved in the early phase of the neural processing of moral evaluation<sup>7</sup>. In another study with

<sup>1</sup>Institutes of Psychological Sciences, Hangzhou Normal University, Hangzhou, 311121, China. <sup>2</sup>Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Hangzhou, 311121, China. Correspondence and requests for materials should be addressed to Q.Y. (email: [jane.qunyang@gmail.com](mailto:jane.qunyang@gmail.com)) or Y.Z. (email: [zhangye.lucia@gmail.com](mailto:zhangye.lucia@gmail.com))

adult participants, time-locked neural responses when distinguishing between intentional and accidental harmful actions were found to occur as quickly as 62 ms after the stimulus onset, demonstrating unexpectedly fast responses by the brain to processing morally laden information<sup>8</sup>.

In line with the research noted above, the temporal relation between moral information and purely negative emotion processing has been investigated in one study using lateralized readiness potentials (LRPs) as a neural index in a Go/No-Go paradigm. The results revealed that the participants unconsciously prepared to respond to the moral feature of an act with a “left” or “right” hand selection decision before accessing the physical disgust information that enabled the Go/No Go decision, suggesting that the processing of moral information and physical disgust information may occur at different stages, with an intuitively faster response for moral information<sup>9</sup>. This temporal priority of processing moral information over physical disgust was further verified by another study in which moral conditions, regardless of the degree of disgust emotion involved, were found to evoke greater positivity at ~300–400 ms than non-moral conditions; however, disgust conditions with or without moral content were found to evoke larger positive deflections at ~500–600 ms than non-disgust conditions<sup>10</sup>. The results from these two studies indicate that morally offensive stimuli are processed prior to physically disgusting stimuli, supporting the notion that the early processing of moral cognition may reflect a unique intuition about the right or wrong categorization process instead of pure emotional reactions<sup>10</sup>. However, in a lexical decision task, Luo and colleagues demonstrated that core disgust words are distinguished from neutral words in the early EPN component whereas the features of moral words are accessible in subsequent N320 and N400 components<sup>11</sup>. Similar results have been obtained in two recent ERP studies, demonstrating an early disgust effect when processing pictorial stimuli with moral, disgust or neutral content<sup>12,13</sup>.

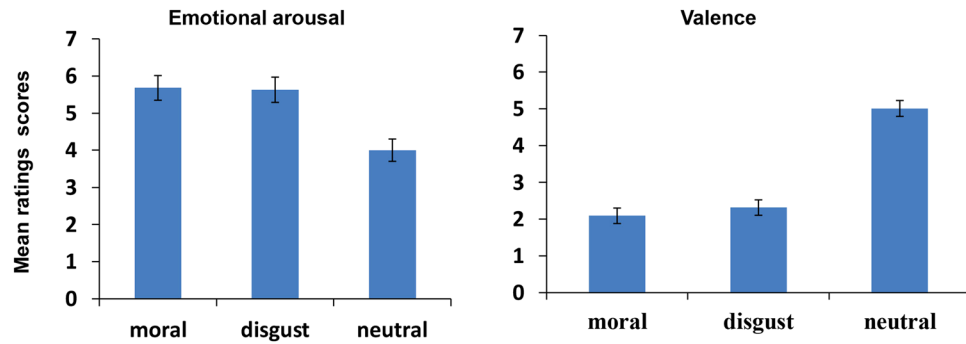
As shown above, there has been growing interest among moral psychologists in clarifying the online processing of moral intuition. However, the evidence to date seems somewhat inconsistent. Various ERP components have been found to be linked to spontaneous moral cognition. Although the time windows for the early stages of moral judgments were approximately ~200–300 ms in some studies<sup>3,4,6,14</sup>, others demonstrated relatively earlier (before 100 ms) or later (after 300 ms) accessibility of automatic moral evaluation in the time-course of processing<sup>5,8,10,11</sup>; even more contradictorily, an early moral effect was detected in some ERP studies prior to a pure emotion effect but was observed after emotional processing in other ERP studies.

In the present study, we attempt to extend the results of previous research by utilizing the recognition potential (RP) as the physiological index to examine how rapidly moral cognition is brought to bear on lexical processing. As indicated by previous research, the RP commonly peaks at approximately ~200–300 ms in the parieto-occipital region in response to meaningful stimuli<sup>11,15</sup>. The RP is typically obtained by the Rapid Stream Stimulation (RSS) paradigm, which is characterized by the alternate presentation of recognizable and non-recognizable stimuli (background) at a high rate without any inter-stimulus intervals. The rapid rate of stimulus presentation in the procedure (normally, the stimulus onset asynchrony [SOA] of each stimulus is 250 ms) is able to force the participants to process the stimuli with their full attention and to attenuate the contamination of irrelevant variables in the electrophysiological signals of interest<sup>16</sup>. It has been suggested that the RP is linked to the pre-semantic processing of words and that it is particularly related to the processing of semantic categorical features<sup>17,18</sup>. Several groups of moral researchers have proposed in their work that the early stage of the temporal dynamics in moral judgment may involve a rapid and intuitive “right-or-wrong” categorization process<sup>5,6,10</sup>. Thus, we expect the RP component to reflect the categorical moral evaluation in early lexical processing.

In addition, we intend to determine the possible reasons that can account for the variability of the early moral effect. Morality is universal but culturally and individually variable<sup>19</sup>. Evidence suggests that individual differences emerge in the distinct psychological processes involved in moral judgment, such as emotional reaction, controlled cognition and mental-state reasoning<sup>20</sup>. Therefore, a plausible method by which to account for the inconsistency between the existing data may be in terms of a subject’s sensitivity to moral values. Individual variation in moral preferences has been reported to be an important predictor of both behavioral and neural responses in social decision-making tasks<sup>21,22</sup>.

Moral foundation theory has developed validated measures of personal moral preferences. According to this theory, five universal moral foundations are included: care/harm (relates to an individual’s sensitivity to feel the suffering of others), fairness/cheating (relates to an individual’s sensitivity to issues of unfair treatment, justice and rights), loyalty/betrayal (relates to concerns for group membership and addresses issues such as loyalty, self-sacrifice and betrayal), authority/subversion (relates to issues of maintaining social order and engaging in hierarchical social interactions) and sanctity/degradation (focuses on concerns about physical and spiritual purity, disgust and contamination). Harm and fairness are categorized as the individualizing foundation because they focus on individualizing moral virtues. The other three are identified as the binding foundation because the underlying moral virtues in each of the foundations can bind groups together for greater strength. The Moral Foundation Questionnaire (MFQ) has been developed as a tool to assess a person’s endorsement of the five foundations<sup>23–25</sup>. Cross-cultural studies have tested the five-factor model of moral foundations in independent populations and provided evidence for the universality of the MFQ<sup>26,27</sup>.

Moral foundations endorsements have been demonstrated to be robustly related to social attitudes and actual behaviors. For example, moral foundation endorsements have been shown to reflect an individual’s political orientation in American culture. In general, political liberals endorse moral values based on the two individualizing foundations, whereas conservatives endorse an additional set of virtues based on the binding foundations<sup>28</sup>. Furthermore, the MFQ is proven to have the ability to predict actual voting behaviors<sup>29</sup>. Additionally, there has been rich evidence indicating that the moral concerns of participants, reflected by MFQ scores, are closely related to individual moral decisions across various contexts. In a recent study, individual differences in multiple moral foundations were found to be predictive of participants’ endorsement of causing harm to one person for the sake of saving more people<sup>30</sup>. In multiple studies designed to examine people’s moral decision-making in video games, the researchers consistently established an association between the



**Figure 1.** Self-reported rating Scores (M+/-SEM) of the emotional arousal and valence for the moral, disgust and neutral words.

participants' MFQ scores and their moral choices in the fictional game world<sup>31–33</sup>. The relative salience of a given moral foundation is related not only to the reasoning process but also to the likelihood of complying with that foundation in their playing behaviors<sup>31,32</sup>. However, to the best of our knowledge, few studies to date have investigated the relationship between the relative salience of moral foundations and the neural responses of moral judgments.

To this end, we measured individual participants' attitudes toward multiple foundational moral values using the MFQ and adopted the RSS procedure to evoke the RP to determine how individual differences in moral preferences are related to the early recognition of moral information in lexical processing. The key prediction is that the degree to which the features of moral information can be attended and recognized at the early semantic processing stage may depend on an individual's sensitivity to certain moral foundations. Specifically, only for those with a higher level of endorsement of universal moral concerns (such as harm, fairness) may the processing of moral information be of more importance with motivational and affective concerns and thus attract more attention, leading to larger RP amplitudes for moral stimuli compared to physically aversive or emotionally neutral stimuli. Furthermore, the two groups with different moral preferences should exhibit different RP amplitudes induced by moral words.

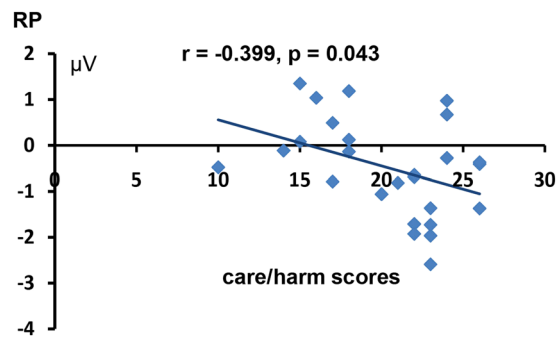
## Results

A one-way repeated-measures analysis of variance (ANOVA) with the stimulus types as the within-subject variable was conducted on the rating scores of emotional arousal and valence by the 27 participants from the formal ERP experiment (see Fig. 1). The arousal results showed that there was a significant main effect of the stimulus types ( $F(2, 52) = 10.982, p = 0.002, \eta_p^2 = 0.297$ ). The moral and disgust words were rated as significantly more emotionally arousing than the neutral words (Bonferroni correction,  $p = 0.004$ ). No differences between the moral words and the disgust words were found. The valence results showed that there was a significant main effect of the stimulus types ( $F(2, 52) = 110.990, p = 0.000, \eta_p^2 = 0.810$ ). The moral words and the disgust words showed no significant differences in valence, but both were rated as remarkably less pleasant than the neutral words (Bonferroni correction,  $p < 0.001$ ).

A two-way repeated measures ANOVA with the stimulus types and electrodes as within-subject variables conducted on the mean amplitudes of the RP revealed a significant main effect of the stimulus types ( $F(2, 52) = 4.503, p = 0.023, \eta_p^2 = 0.148$ ) and a marginally significant main effect of the electrodes and interaction effect between the stimulus types and electrodes ( $F(3, 78) = 3.284, p = 0.056, \eta_p^2 = 0.112$ ;  $F(6, 156) = 2.146, p = 0.066, \eta_p^2 = 0.076$ ). The Moral words and the disgust words evoked significantly larger RPs than the neutral words (Bonferroni correction,  $p = 0.07$ ), but did not differentiate from each other.

However, an evident individual variability of the grand-averaged ERP waveforms was observed. Specifically, some participants displayed larger RPs for the moral words relative to the other two types of words, whereas others displayed larger RPs for the disgust words. To further reveal the inter-individual differences of the evoked RPs among all participants, we first computed the net moral or disgust RP effect by subtracting the averaged RP amplitudes of the neutral words across all electrodes from those of the moral words or the disgust words for each participant. Then, we examined the correlations between the net moral or disgust RP effect and their scores on the MFQ-30 and Disgust Scale. The data for one of the participants were excluded from the correlation and the following analysis because the moral RP difference wave amplitudes of this participant went beyond 2.5 standard deviations of the mean number. As shown in Fig. 2, the net moral RP effect was found to be negatively correlated with the Harm/Care foundation scores of the MFQ-30 ( $r = -0.399, p = 0.043$ ). However, the correlation failed to reach significance when the conservative Bonferroni correction method was applied. We did not observe any other significant correlations.

To further reveal the relation between the participants' endorsements of moral foundations and the early neural processing of moral information, we used a median split on the Harm/Care scores of the MFQ-30 and assigned the participants who scored above the median score ( $M = 22$ ) to the high-sensitivity group ( $N = 14$ ) and the participants who scored below to the low-sensitivity group ( $N = 12$ ). The scores on each subscale of Moral Foundation Questionnaire and the Disgust Scale for the two groups are presented in Table 1.



**Figure 2.** Correlation between the net moral RP effect and the harm/care scores of the MFQ-30.

	High-sensitivity group	Low-sensitivity group
Total Scores(Disgust Scale)	68.979(10.850)	63.000(12.030)
Pathogen-related	14.929(4.287)	15.0833(4.078)
Body envelope	18.643 (4.781)	16.667(6.652)
Moral	19.357(4.253)	17.083(3.777)
Sex-related	15.857(4.833)	14.167(5.357)
Total Scores(MFQ-30)	95.714(7.956)	83.500(13.028)
Harm/Care	23.571(1.505)	16.583(2.906)
Fairness/Reciprocity	20.857(3.527)	19.250(3.251)
Ingroup/Loyalty	21.571(2.901)	18.917(5.265)
Authority/Respect	18.429(3.204)	17.083 (3.343)
Purity/Sanctity	16.857 (2.797)	16.500(3.371)

**Table 1.** Scores on the Moral Foundation Questionnaire and the Disgust Scale (mean and stand deviation) for the high-sensitivity group and the low-sensitivity group.

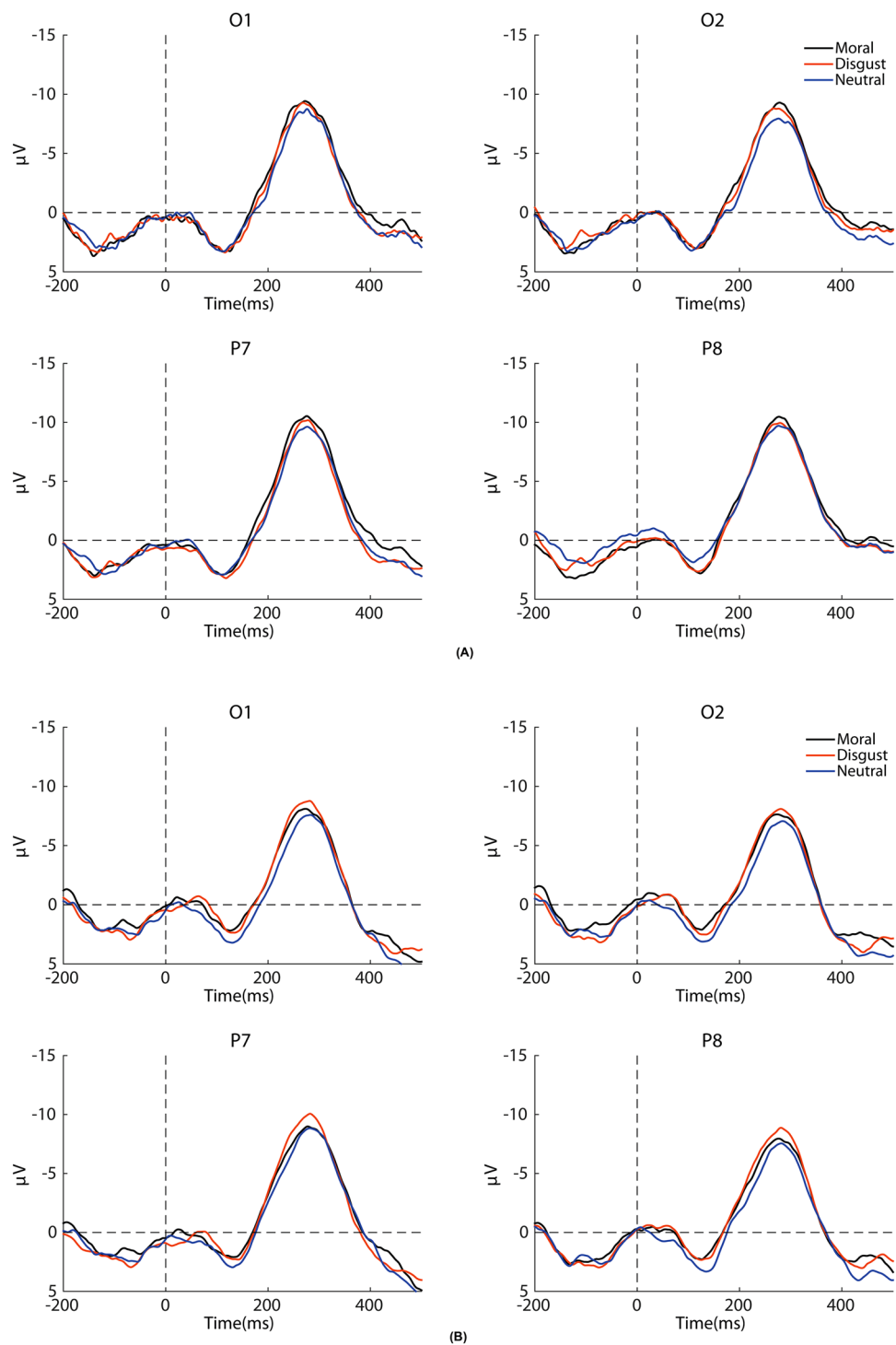
We then conducted a two-way mixed ANOVA with the stimulus types as a within-subject factor and the group as a between-subject factor on the mean amplitudes of the RP across all electrodes. The main effect of the stimulus types and the interaction between the two independent variables was significant ( $F(2, 48) = 5.667, p = 0.011, \eta_p^2 = 0.191$ ;  $F(2, 48) = 5.430, p = 0.013, \eta_p^2 = 0.184$ ). The simple effects analysis of the group and the stimulus type interaction was subsequently performed: for the high-sensitivity group, the moral words evoked significantly larger RPs than the neutral words (Bonferroni correction,  $p = 0.002$ ) (see Fig. 3A). No differences between the disgust words and the neutral words were observed; for the low-sensitivity group, the disgust words evoked significantly larger RPs than the neutral words (Bonferroni correction,  $p = 0.016$ ). No differences between the moral words and the neutral words were found (see Fig. 3B).

Subsequently, we compared the net moral or disgust RP effect between the two groups. The high-sensitivity group demonstrated a significantly larger net moral effect than the low-sensitivity group ( $t = -2.800, p = 0.01$ ). However, no significant differences for the net disgust effect between the two groups were observed ( $t = 1.212, p > 0.05$ ) (see Fig. 4).

## Discussion

The present study used the RSS procedure to evoke the RP in an attempt to examine the neural dynamics of moral cognition in the early lexical processing stage, with a special interest in determining whether the early recognition of immorality depends on individual differences in moral preferences. The main results showed that the moral words induced larger RP amplitudes compared to the neutral words. However, this was only true for those participants who were more sensitive to the harm/care moral foundation. The two groups with different levels of endorsement of the harm/care foundation showed distinct early moral effects during the pre-semantic lexical processing.

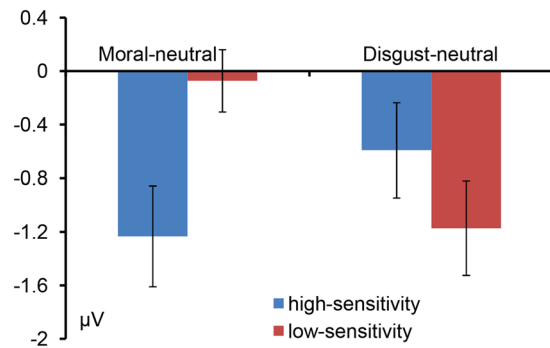
The RP was initially reported to reflect visual processing based on individual learning experiences<sup>34</sup>. Growing evidence then emerged to support the view that the RP is linked to the pre-semantic processing of images or words. It has consistently been shown that the RP arose in the lower level of word analysis during the reading process whereas its amplitude progressively increased by means of appropriate semantic processing<sup>17, 35</sup>. Moreover, RP amplitudes vary as a function of the semantic categories; for example, the names of animals have been found to display larger RP amplitudes than those of non-animals. As an index of rudimentary semantic categorization, the RP has been verified by adequate evidence<sup>17, 18, 35, 36</sup>. It has been further demonstrated that the RP is strongly influenced by the allocation of attentional resources, with amplified RP in response to attended relative to unattended recognizable stimuli<sup>16</sup>. In addition, emotional states have also been found to be able to modulate the



**Figure 3.** Grand average waveforms for the RP component at O1, O2, P7, and P8 across the three types of words for the high-sensitivity group (A) and the low-sensitivity group (B).

amplitudes of the RP component, with slightly negative emotional states evoking larger RPs than highly negative emotional states<sup>37</sup>.

According to our study, the detection of moral information occurred as early as the pre-semantic processing stage when the participants were only allowed to have a short period of time of 250 ms to recognize each word. As shown from the behavioral results on the emotional arousal and valence, the moral words and the disgust words were rated as equally unpleasant and arousing. However, only the RP amplitudes induced by the moral words were found to be dissociated from those induced by the neutral words in the high-sensitivity group, indicating that the early moral effect cannot simply be attributed to the differences in the emotional states. Instead, we assume that the unique features of immorality spontaneously capture selective attention and hence modulate



**Figure 4.** The mean amplitudes of the moral-neutral difference wave for the RP (the net moral effect) and the mean amplitudes of the disgust-neutral difference wave (the net disgust effect) for each group of participants.

the early lexical-semantic processing indexed by the RP. More importantly, we find that the early net moral RP effect was related to the harm/care foundation but not to disgust sensitivity, suggesting that the RP induced by the moral words in this study may reflect the spontaneous evaluative processing of the moral status of an act rather than some general affective reaction.

The results from our study add new evidence for the notion that early ERP components are associated with the rapid information processing involved in moral judgment<sup>3,4,7–9</sup>. Previous results related to the early stage of moral judgment have been mixed. Similar to our findings, a previous moral study using linguistic materials showed that a word indicating value-based disagreement elicited early neural responses between 200 and 250 ms, revealing the brain's fast affective responses to morally wrong statements<sup>14</sup>. Other studies using pictorial visual stimuli revealed that early negativities such as N2 and EPN are sensitive to the automatic differentiation between positive and negative moral scenarios<sup>6,7</sup>. In particular, it has been believed that the EPN effect reflects mechanisms that are similar to the RP component<sup>38</sup>. However, the spontaneous moral evaluation effect has not been reliably detected in early ERP components. Instead, moral information is not accessible until 300 ms after stimulus onset, as demonstrated by some findings<sup>5,10,11</sup>. These inconsistencies may be due to several factors, such as the differences in the stimuli or experimental paradigms used in previous research. Among others, the individual differences in moral preferences particularly warrant observation.

As shown by our data, the early detection of immoral words only occurred with some participants. The variation in the early moral effect manifested in the RP can largely be explained by the individual differences in moral preferences in the harm and care foundation. On the one hand, larger RPs were obtained when viewing the moral words compared to the neutral words only for those who had greater moral concerns about the harm/care foundation. On the other hand, those participants who were more sensitive to the pain of others displayed a larger net moral effect than those who were less sensitive. The harm/care foundation is believed to be associated with the ability to feel the pain of others and the virtues of kindness and gentleness<sup>39</sup>. Harm and fairness have been regarded as the core values in moral concerns in traditional moral psychology<sup>24</sup>. Recently, some researchers have gone further to argue for the central role of harm in evaluating immoral behaviors, asserting that perceived harm drives moral judgment<sup>40</sup>. In this study, the sensitivity to the harm and care foundation, we believe, may determine the relative adaptive importance of moral information in the lexical processing of words, which in turn may lead to the varied allocation of “motivated” attention in the early neural responses.

Previously, several ERP studies have documented the relation between the neural correlates of moral evaluation and individual dispositional factors. As indicated in the study by Chiu Loke *et al.* the P3 component, which reflects the participants' moral reasoning about prosocial-helping behaviors, is related to their self-rating scores on a prosocial personality questionnaire. The more prosocial an individual was, the longer the P3 peak latency that the researchers observed<sup>41</sup>. Similarly, Yoder *et al.* found that the differential amplitudes of late positivities (~300–600 ms) when participants made moral judgments about good and bad behaviors were significantly correlated with individual dispositions in cognitive empathy<sup>6</sup>. However, the ERP-disposition relationship in both of these studies was only observed in late time windows. To the best of our knowledge, the present study is the first to reveal that the individual differences in the harm and care concern are related to the neural processing of moral content in such early ERP components as the RP.

Future research is needed to elucidate the discrepancy in the brain-disposition results between our study and others. One noteworthy factor regarding the interpretations is cultural differences in moral judgment<sup>42</sup>. Given the vast body of research in moral neuroscience, it is surprising that only a few studies have broached the topic of how cultural factors influence the neural processing of moral judgment. A functional magnetic resonance imaging (fMRI) study compared the neural correlates of moral decision-making between Korean and American subjects. Enhanced brain activities in the brain regions were shown to be associated with intuitive judgments for Koreans, whereas they were shown in those brain areas with cognitive control processes for Americans in the moral-personal condition<sup>43</sup>. Another piece of evidence using the ERP method detected neural distinction between personal and impersonal dilemmas in the P3 component between ~280–380 ms for Westerners whereas it occurred in the P260 component between ~200–300 ms for Chinese when all participants needed to make judgments about the appropriateness of an action in hypothetical moral dilemmas<sup>44</sup>. Therefore, it is possible that the present findings may be specific to the population in a certain culture. By all accounts, considering cultural

variables that affect the neural processing of moral judgment may be useful to resolve the inconsistencies that have emerged in previous research.

Interestingly, the dissociation between the disgust words and the neutral words was also reflected by the RP in some of the participants. Participants who had lower scores in the harm and care foundation showed significantly larger RP amplitudes for the disgust words than for the neutral words. However, the correlation between the mean RP amplitudes of the disgust-neutral difference wave and the scores on the disgust subscale was not significant. Moreover, we did not find any difference in the net disgust RP effect between the two groups. In fact, we found that the RP amplitudes induced by the disgust words were not much different between the two groups (high-sensitivity group: mean =  $-9.614$ ; low-sensitivity group: mean =  $-9.875$ ; no significant differences). Perhaps this result is not only because the participants in the group were slightly more alert to the disgust words but also because they responded with slightly less sensitivity to the neutral words (high-sensitivity group: mean =  $-9.074$ ; low-sensitivity group: mean =  $-8.702$ ; no significant differences). At present, however, we can hardly draw any conclusions about the underlying cause of the dissociation between the disgust words and the neutral words.

Nonetheless, what we found in the current research promotes our understanding of the relationship between the processing of moral judgment and that of the emotion of disgust. Behaviors that violate moral rules have been documented to be one of the primary elicitors of disgust<sup>45</sup>. It has been proposed that disgust elicitors have expanded from the physical domain to the socio-moral domain<sup>45,46</sup>. Not only is moral disgust linguistically analogous to physical disgust, but it may also represent a sub-set of the experience of disgust<sup>47</sup>. In this study, we found that moral and disgust stimuli showed different time-locked neural response patterns in the two groups. Moreover, the moral effect in the RP component was not correlated with behavioral ratings on the disgust sensitivity scale. Along with our previous studies, we consistently support the notion that the evaluation of immoral information is biologically distinct from that of physically disgusting information in the temporal dimension<sup>9,10</sup>.

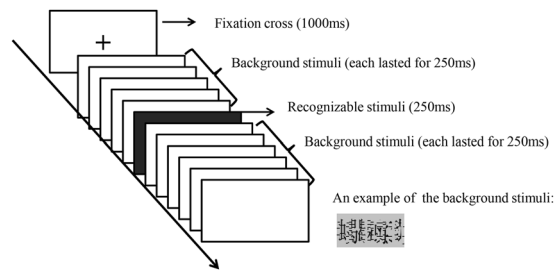
**Summary and Limitations.** Collectively, we demonstrated that the moral features of lexical words are accessible at the pre-semantic processing stage with individual differences. The early recognition of moral information in the time course hinges on individuals' moral concerns about harm and care. Only those participants who had a high sensitivity to the suffering of others displayed the ability to distinguish between moral and neutral stimuli in the RP component. Despite these valuable findings, some limitations of this study must be addressed. First, although the moral and disgust stimuli were matched in emotional arousal and valence, the moral words may most likely be considered disgusting by the participants. As argued by previous research, the emotion of disgust can be induced by harmful physical substances and by immoral behaviors<sup>47,48</sup>. Evidence from previous research and the present study that moral stimuli, regardless of the level of the emotion of disgust, can be temporally dissociable from purely disgusting stimuli relieves some of the concern over this issue<sup>9,10</sup>. Nonetheless, the failure to separate the disgusting feature from the moral stimuli created a potential confounding factor that may have biased the results. Another concern with the experimental materials is that the moral words that we used were more representative of the harm dimension. Therefore, how the other foundations would be related to the RP if more types of moral words were adopted in the experiment remains inconclusive. Second, despite a strong numerical trend, as expected, the correlation between the harm/care concern and the RP amplitudes of the moral items failed to reach significance when a conservative correction method was applied. Meanwhile, the median split method is normally not a favorable technique for dividing the participants based on some dispositional ratings. In the future, it should be beneficial to use a larger sample to examine the brain-moral dispositions relation. Recruiting more participants, we can screen out two groups with larger moral dispositional differences to replicate the observed results of this study. Finally, due to the characteristics of the experimental paradigm, we were limited to examining the temporal dynamics of the moral judgment in the early components, which made it difficult to compare the findings with those from previous studies since they found the association between brain activities and moral dispositions in later components.

These issues notwithstanding, the results of this research provide new evidence for the fast and intuitive processing of moral information in the early pre-semantic stage. Moreover, the study offers new insights into the causes of the variation in the early moral effect in research.

## Methods

**Participants.** Thirty adults were recruited into the study. Three subjects were excluded from further analysis because, for each of them, more than 50% of the experimental trials were removed as a result of artifacts in the data preprocessing (see the Method section for the criteria for artifact removal). Ultimately, twenty-seven adults (17 females) were included in the ERP data analysis. The age of the participants ranged from 18 to 25 years old ( $M = 21.48$ ,  $SD = 2.62$ ). All participants were right-handed and had normal or corrected-to-normal vision. No brain damage or neurological or psychological diseases were reported. The experimental procedure of this study was approved by the Institutional Review Board of Hangzhou Normal University, and all methods conformed to the relevant guidelines and regulations with regard to the use of human participants. Each participant signed a written, informed consent form prior to the ERP experiment and was compensated RMB¥ 50 after the experiment.

**Stimuli.** The entire set of stimuli consisted of recognizable words and non-recognizable background materials. Four types of words were included as the recognizable materials. Each of the four types of words was composed of three characters. The moral words depicted morally wrong behaviors (e.g., eating a person—吃活人); the disgust words depicted morally acceptable but physically disgusting behaviors (e.g., eating rotten meat—吃腐肉); the neutral words depicted both morally and emotionally neutral behaviors (e.g., eating spaghetti—吃意面); the target words depicted the name of a city or country (e.g., Hei Long Jiang—黑龙江). For each set of moral, disgust



**Figure 5.** An illustration of the RSS procedure for the stimulus presentation.

and neutral words, the first character was the same as a verb, denoting an action, and the last two characters combined as a noun, denoting an object.

The moral, disgust and neutral words used in the formal experiment were carefully selected based on two rounds of pilot studies involving 36 participants (15 males, 21 females). First, the participants were asked to describe whether they think each behavior is morally wrong or whether it is exclusively disgusting. A word was categorized as a moral word if 80% or more of the participants judged a behavior that the word depicted to be morally wrong. A word was categorized as a disgust word if 80% or more of the participants judged the behavior to be exclusively physically disgusting. A word was categorized as a neutral word if 80% or more of the participants judged the behavior to be neither morally wrong nor physically disgusting. Second, the participants were instructed to assess emotional arousal (from very calm to very exciting) and valence (from very unpleasant to very pleasant) using a Likert scale ranging from 1 to 9. Those words having arousal ratings lower than 5 and valence ratings higher than 5 in the moral and disgust categories were excluded. Those words having arousal ratings lower than 5 and valence ratings between 4 and 6 were categorized as neutral words. Moreover, emotional arousal and valence was matched between the moral words and the disgust words. Ultimately, the moral, disgust and neutral categories were composed of 38 words, and another 20 words were included as the target stimuli.

Based on previous RP research, the unrecognizable background stimuli were composed of the recognizable words, which were cut into fragments and restructured using a customized MATLAB script. These background stimuli have no set structures or meanings but were matched in visual attributes with the four types of recognizable stimuli.

Subsequently, 134 recognizable and the corresponding 134 background stimuli constituted the formal set of stimuli for the ERP experiment.

**Procedure.** The experiment was conducted in a semi-dark, well shielded and quiet room. The participants were seated at a viewing distance of approximately 60 cm from the computer screen and were instructed to avoid blinking their eyes or moving their bodies as much as possible while keeping their eyes fixated at the center of the computer screen. All stimuli were presented in the center of a 17-inch computer screen using the Song font style with a 28-point font size. The task of this experiment was to respond to the target words. Both the reaction time and accuracy of the behavioral responses were recorded.

At the beginning of a trial, a “fixation cross” was presented for 1000 ms to remind the participants to keep their eyes focused on the center of the screen. Subsequently, 3 to 5 background stimuli were presented. Then, a recognizable word appeared, which was followed by another 4 to 6 background stimuli. Each stimulus lasted for 250 ms (see Fig. 5). The participants were instructed to press the “1” key when they detected the name of a city or a country, namely, the target word. The design of the task is, in fact, irrelevant to the true purpose of the experiment. The implicit task is commonly used in the RSS paradigm<sup>49,50</sup>. Meaningful target and non-target words were shown to evoke a “fairly equal RP amplitude and same latency”<sup>50</sup>. The recognizable stimuli including target words were presented randomly during the experiment; thus, the participants had to keep their attention on all recognizable stimuli and extract meaning from each recognizable stimulus. Each stimulus of the moral, disgust and neutral words was repeated three times at most. The target words were repeated twice at most. The neighboring background materials within a trial were deliberately made different to prevent visual adaptation.

Prior to the formal experiment, 10 practice trials were arranged to familiarize the participants with the experimental procedure. The participants were only allowed to start the formal experiment when their accuracy rate reached more than 90% in the practice session. The formal test consisted of six blocks of 46 trials each. The participants were allowed to take a short break after each block.

After the ERP experiment, individuals’ moral preferences were measured by the 30-item Moral Foundations Questionnaire (MFQ-30) which is composed of five dimensions: harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect and purity/sanctity. The participants were asked to rate both the moral relevance of 15 foundation-related considerations (e.g., whether someone suffered emotionally—harm/care) and the level of agreement with 15 moral judgment statements (e.g., compassion for those who are suffering is the most crucial virtue—harm/care). The scale has been proven to have fairly acceptable reliability and validity<sup>51,52</sup>. In addition, to exclude the possibility that the early recognition of moral information could be simply related to a person’s pure emotional sensitivity, the participants were also required to complete the disgust sensitivity scale because disgust is the emotion that is most associated with moral appraisal<sup>53–57</sup>. Disgust sensitivity was assessed using the Chinese Version of the Disgust Scale, which was standardized and validated among Chinese populations by our group on



the basis of the Disgust Scale developed by Haidt *et al.*<sup>58</sup> and the Three Domains Disgust Scale developed by Tybur *et al.*<sup>59</sup>. Four dimensions were included in the Disgust Scale: pathogen-related disgust, body envelope disgust, sex-related disgust and moral disgust. This measure has been shown to have a Cronbach's  $\alpha$  coefficient of 0.92 for the entire scale and values higher than 0.7 for each sub-dimension.

At the end of the experiment, each participant was instructed to rate the arousal and valence for each moral, disgust and neutral word on a scale ranging from 1 to 9.

**Electrophysiological recording and analysis.** EEG data were acquired from 32 channels using tin electrodes mounted in an elastic cap (Brain Products GmbH, Munich, Germany). The vertical electrooculogram (VEOG) was recorded with an electrode placed below the left eye, and the horizontal electrooculogram (HEOG) was recorded with an electrode placed next to the orbital rim of the right eye. All inter-electrode impedances were maintained below 5 k $\Omega$ . The EEG and EOG signals were amplified using a DC 0.016–100 Hz band-pass and were recorded with a sample rate of 1000 Hz. The EEG data were analyzed offline using EEGlab<sup>60</sup>. The data were filtered using a 0.01–40 Hz band-pass infinite impulse response (IIR) filter. Trials with mean EOG voltage exceeding  $\pm 75 \mu\text{V}$  and those contaminated by artifacts as a result of amplifier clipping bursts of electromyographic activity or peak-to-peak deflection exceeding  $\pm 75 \mu\text{V}$  were excluded from further data analysis. ERP averages were computed offline for each type of words; the data were time-locked to the onset of the recognizable words.

Only ERPs elicited by the moral, disgust and neutral words were of interest to us and were thus further analyzed. The mean numbers of remaining trials among all participants for further analysis after data preprocessing were 64 (accounting for 81% of the total trials), 65 (82%) and 63 (80%) for the moral, disgust and neutral conditions, respectively. The averaged epochs for ERPs were 700 ms, which included a 200 ms pre-stimulus and 500 ms post-stimulus activity. Based on the observation of grand-averaged ERP waveforms, a negative component in the parieto-occipital region peaking at approximately 275 ms was elicited by the different types of words. The parieto-occipital distribution is consistent with previous observations<sup>15,50</sup>. Therefore, the following 4 electrode sites were selected for statistical analysis: O1, O2, P7, and P8. We measured the RP of each individual between 160 and 360 ms after recognizable stimulus onset, in which the RP activity was prominent on average. The p-value was corrected for deviations according to Greenhouse Geisser for all analyses.

## References

- Kohlberg, L. In *Handbook of socialization theory and research* (ed. Goslin, D. A.) 151–235 (Academic Press, 1969).
- Waldmann, M. R., Nagel, J. & Wiegmann, A. In *The Oxford Handbook of Thinking and Reasoning* (eds Holyoak, K. J. & Morrison, R. G.) 364–389 (Oxford University Press, 2012).
- Chen, P., Qiu, J., Li, H. & Zhang, Q. Spatiotemporal cortical activation underlying dilemma decision-making: an event-related potential study. *Biol. Psychol.* **82**, 111–115 (2009).
- Sarlo, M. *et al.* Temporal Dynamics of Cognitive-Emotional Interplay in Moral Decision-making. *J. Cogn. Neurosci.* **24**, 1018–1029 (2011).
- Leuthold, H., Kunkel, A., Mackenzie, I. G. & Filik, R. Online processing of moral transgressions: ERP evidence for spontaneous evaluation. *Soc. Cogn. and Affect. Neurosci.* **10**, 1021–1029 (2015).
- Yoder, K. J. & Decety, J. Spatiotemporal neural dynamics of moral judgment: A high-density ERP study. *Neuropsychologia* **60**, 39–45 (2014).
- Cowell, J. M. & Decety, J. The neuroscience of implicit moral evaluation and its relation to generosity in early childhood. *Curr. Biol.* **25**, 93–97 (2015).
- Decety, J. & Cacioppo, S. The speed of morality: a high-density electrical neuroimaging study. *J. Neurophysiol.* **108**, 3068–3072 (2012).
- Yang, Q. *et al.* Temporal Dynamics of Disgust and Morality: An Event-Related Potential Study. *PLoS ONE* **8**, e65094 (2013).
- Yang, Q., Li, A., Xiao, X., Zhang, Y. & Tian, X. Dissociation between morality and disgust: An event-related potential study. *Int. J. Psychophysiol.* **94**, 84–91 (2014).
- Luo, Y. *et al.* Core disgust and moral disgust are related to distinct spatiotemporal patterns of neural processing: an event-related potential study. *Biol. Psychol.* **94**, 242–248 (2013).
- Ding, D., Zhang, X. & Chen, X. The ERPs Study on the Divergent Effect of Processing Core and Moral Disgust Stimuli at Different Level of Attentional Resources. *Psychol. Sci.* **39**, 265–271 (2016).
- Zhang, X., Guo, Q., Zhang, Y., Lou, L. & Ding, D. Different timing features in brain processing of core and moral disgust pictures: an event-related potentials study. *PLoS One* **10**, e0128531 (2015).
- Van Berkum, J. J., Holleman, B., Nieuwland, M., Otten, M. & Murre, J. Right or wrong? The brain's fast response to morally objectionable statements. *PsycholSci* **20**, 1092–1099 (2009).
- Zhang, Y., Yuan, J., Bao, B. & Zhang, Q. The recognition potential and rotated Chinese characters. *Brain Res.* **1233**, 98–105 (2008).
- Martin-Loeches, M. The gate for reading: reflections on the recognition potential. *Brain res. rev.* **53**, 89–97 (2007).
- Martin-Loeches, M., Hinojosa, J. A., Gómez-Jarabo, G. & Rubia, F. J. An early electrophysiological sign of semantic processing in basal extrastriate areas. *Psychophysiology* **38**, 114–124 (2001).
- Martin-Loeches, M., Muñoz, F., Casado, P., Hinojosa, J. A. & Molina, V. An electrophysiological (ERP) component, the recognition potential, in the assessment of brain semantic networks in patients with schizophrenia. *Schizophr. Res.* **71**, 393–404 (2004).
- Haidt, J. The new synthesis in moral psychology. *Science* **316**, 998–1002 (2007).
- Young, L. & Saxe, R. Moral Universals and Individual Differences. *Emot. Rev.* **3**, 323–324 (2011).
- Palmer, C. J. *et al.* Individual Differences in Moral Behaviour: A Role for Response to Risk and Uncertainty? *Neuroethics* **6**, 97–103 (2012).
- Prehn, K. *et al.* Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. *Soc. Cogn. Affect. Neurosci.* **3**, 33–46 (2008).
- Clifford, S., Iyengar, V., Cabeza, R. & Sinnott-Armstrong, W. Moral foundations vignettes: a standardized stimulus database of scenarios based on moral foundations theory. *Behav. res. meth.* **47**, 1178–1198 (2015).
- Glover, R. J. *et al.* Moral rationality and intuition: An exploration of relationships between the Defining Issues Test and the Moral Foundations Questionnaire. *J. Moral Educ.* **43**, 395–412 (2014).
- Vaisey, S. & Miles, A. Tools from moral psychology for measuring personal moral culture. *Theor. Soc.* **43**, 311–332 (2014).
- Davies, C. L., Sibley, C. G. & Liu, J. H. Confirmatory Factor Analysis of the Moral Foundations Questionnaire Independent Scale Validation in a New Zealand Sample. *Soc. Psychol.-Germany* **45**, 431–436 (2014).

27. Yilmaz, O., Harma, M., Bahcekapili, H. G. & Cesur, S. Validation of the Moral Foundations Questionnaire in Turkey and its relation to cultural schemas of individualism and collectivism. *Personal. Individ. Diff.* **99**, 149–154 (2016).
28. Haidt, J. & Graham, J. When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals may not Recognize. *Soc. Just. Res.* **20**, 98–116 (2007).
29. Franks, A. S. & Scherr, K. C. Using Moral Foundations to Predict Voting Behavior: Regression Models from the 2012 US. Presidential Election. *Anal. Soc. Issues Public Policy* **15**, 213–232 (2015).
30. Crone, D. L. & Laham, S. M. Multiple moral foundations predict responses to sacrificial dilemmas. *Personal. Individ. Diff.* **85**, 60–65 (2015).
31. Joeckel, S., Bowman, N. D. & Dogruel, L. Gut or Game? The Influence of Moral Intuitions on Decisions in Video Games. *Media Psychol.* **15**, 460–485 (2012).
32. Krcmar, M. & Cingel, D. P. Moral Foundations Theory and Moral Reasoning in Video Game Play: Using Real-Life Morality in a Game Context. *J. Broadcast. Electron. Media* **60**, 87–103 (2016).
33. Weaver, A. J. & Lewis, N. Mirrored Morality: An Exploration of Moral Choice in Video Games. *Cyberpsychol. Behav. Soc. Netw.* **15**, 610–614 (2012).
34. Rudell, A. P. Rapid Stream Stimulation and the Recognition Potential. *Electroenceph. Clin. Neurophysiol.* **83**, 77–82 (1992).
35. Martin-Loeches, M., Hinojosa, J. A., Fernández-Frias, C. & Rubia, F. J. Functional differences in the semantic processing of concrete and abstract words. *Neuropsychologia* **39**, 1086–1096 (2001).
36. Hinojosa, J. A. *et al.* Semantic processing of open and closed-class words: an event-related potentials study. *Cogn. Brain Res.* **11**, 397–407 (2001).
37. Guo, T. M., Chen, M. & Peng, D. Emotional States Modulate the Recognition Potential during Word Processing. *PLoS ONE* **7** (2012).
38. Herbert, C., Junghofer, M. & Kissler, J. Event related potentials to emotional adjectives during reading. *Psychophysiology* **45**, 12 (2007).
39. Graham, J. *et al.* Moral foundations theory: The pragmatic validity of moral pluralism. *Adv. Exp. Soc. Psychol.* **47**, 55–130 (2013).
40. Gray, K. Harm concerns predict moral judgments of suicide: Comment on Rottman, Kelemen and Young (2014). *Cognition* **133**, 329–331 (2014).
41. Chiu Loke, I., Evans, A. D. & Lee, K. The neural correlates of reasoning about prosocial-helping decisions: an event-related brain potentials study. *Brain Res.* **1369**, 140–148 (2011).
42. Graham, J., Meindl, P., Beall, E., Johnson, K. M. & Zhang, L. Cultural differences in moral judgment and behavior, across and within societies. *Curr. Opin. Psychol.* **8**, 125–130 (2016).
43. Han, H., Glover, G. H. & Jeong, C. Cultural influences on the neural correlate of moral decision making processes. *Behav. Brain Res.* **259**, 215–228 (2014).
44. Wang, Y., Deng, Y., Sui, D. & Tang, Y. Y. Neural correlates of cultural differences in moral decision making: a combined ERP and sLORETA study. *Neuroreport* **25**, 110–116 (2014).
45. Rozin, P., Haidt, J. & McCauley, C. R. In *Handbook of emotions* (eds Lewis, J. M., Haviland-Jones & Barrett, L. F.) 757–776 (Guilford Press, 2008).
46. Haidt, J., Rozin, P., McCauley, C. R. & Imada, S. Body, psyche, and culture: The relationship between disgust and morality. *Psychol. Dev. Soc.* 107–131 (1997).
47. Chapman, H. A. & Anderson, A. K. Understanding disgust. *Ann. N.Y. Acad. Sci.* **1251**, 62–76 (2012).
48. Rozin, P., Haidt, J. & Fincher, K. From Oral to Moral. *Science* **323**, 1179–1180 (2009).
49. Zhang, Y., Liu, Q. & Zhang, Q. The recognition potential reflects an intermediate level of visual representation. *Neurosci. Lett.* **454**, 86–90 (2009).
50. Martin-Loeches, M., Hinojosa, J. A., Gomez-Jarabo, G. & Rubia, F. J. The recognition potential: An ERP index of lexical access. *Brain lang.* **70**, 364–384 (1999).
51. Graham, J. *et al.* Mapping the moral domain. *J. Pers. Soc. Psychol.* **101**, 366–385 (2011).
52. Haidt, J. & Joseph, C. In *The Innate Mind Vol. 3* (eds Carruthers, P., Laurence, S. & Stich, S.) 367–391 (Oxford, 2007).
53. Eskine, K. J., Kacinik, N. A. & Prinz, J. J. A bad taste in the mouth: gustatory disgust influences moral judgment. *Psychol. Sci.* **22**, 295–299 (2011).
54. Horberg, E. J., Oveis, C., Keltner, D. & Cohen, A. B. Disgust and the moralization of purity. *J. Pers. Soc. Psychol.* **97**, 963–976 (2009).
55. Schnall, S., Haidt, J., Clore, G. L. & Jordan, A. H. Disgust as embodied moral judgment. *Pers. Soc. Psychol. Bull.* **34**, 1096–1109 (2008).
56. Shweder, R. A., Mahapatra, M. & Miller, J. In *The emergence of morality in young children* (eds Kagan, J. & Lamb, S.) 1–83 (University of Chicago Press, 1987).
57. Wheatley, T. & Haidt, J. Hypnotic disgust makes moral judgments more severe. *Psychol. Sci.* **16**, 780–784 (2005).
58. Haidt, J., McCauley, C. & Rozin, P. Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personal. Individ. Diff.* **16**, 701–713 (1994).
59. Tybur, J. M., Lieberman, D. & Griskevicius, V. Microbes, mating, and morality: individual differences in three functional domains of disgust. *J. Pers. Soc. Psychol.* **97**, 103–122 (2009).
60. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).

## Acknowledgements

This research was supported by National Natural Science Foundation of China (Youth Project: 31500907 and 31400974) grants to Qun Yang and Ye Zhang, respectively. We would like to thank Duan Qing fang for her help in the data collection.

## Author Contributions

Y.Z. and Q.Y. contributed equally to the design of the study. Q.Y. prepared the experimental materials, collected the data, analyzed the results and wrote the paper. Y.Z. contributed to the experimental materials and participated in part of the data analysis. L.C. and Y.Z. programmed the experiment, collected the data and preprocessed the ERP data. All three authors reviewed the manuscript.

## Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017