

Manuscript Number: NLM-17-93R1

Title: Aversive smell associations shape social judgment

Article Type: Research paper

Keywords: Attribution; olfactory conditioning; smell, ventromedial prefrontal cortex; social decision-making

Corresponding Author: Dr. Daniela Schiller,

Corresponding Author's Institution: Icahn School of Medicine at Mount Sinai

First Author: Philipp Homan, M.D., Ph.D.

Order of Authors: Philipp Homan, M.D., Ph.D.; Benjamin A Ely; May Yuan; Tobias Brosch, Ph.D.; John Ng, Ph.D.; Yaacov Trope, Ph.D.; Daniela Schiller, Ph.D.

Abstract: Once associating another person with an unpleasant smell, how do we perceive and judge this person from that moment on? Here, we used aversive olfactory conditioning followed by a social attribution task during functional magnetic resonance imaging to address this question. After conditioning, where one of two faces was repeatedly paired with an aversive smell, the participants reported negative affect when viewing the smell-conditioned but not the neutral face. When subsequently confronted with the smell-conditioned face (without any smell), the participants tended to judge both positive and negative behaviors as indicative of personality traits rather than related to the situation. This effect was predicted by the degree of the preceding olfactory evaluative conditioning. Whole brain analysis of stimulus by stage interaction, indicated differential activation of the ventromedial prefrontal cortex and right angular gyrus to the conditioned versus the neutral person during the attribution phase only. These results suggest that negative smell associations do not simply induce a negative perception of the target person but rather bias the attribution style toward trait attributions. The fact that this bias was evident regardless of behavior valence suggests it may reflect enhanced psychological distance. Thus, the known observation of social rejection triggered by aversive smell may be driven by a shift in social attribution style.



**Mount
Sinai**

Daniela Schiller, PhD

Associate professor

Department of Psychiatry

Department of Neuroscience

One Gustave L. Levy Place, Box 1230

New York, NY 10029

Tel. (212) 824-8977

Fax. (212) 731-7804

Email: daniela.schiller@mssm.edu

June 19th, 2017

Dear Editor,

Please find below the revised manuscript titled "*Aversive smell associations shape social judgment*", which we submit as research article for publication in *Neurobiology of Learning and Memory*.

Thank you for considering our manuscript and giving us the opportunity to reply to the reviewers' comments. Below we provide point-by-point responses, addressing the reviewers' concerns in full and improving the clarity of the manuscript.

We believe that this work is appropriate for publication in *Neurobiology of Learning and Memory* because of its novelty, broad significance, and interest to researchers across a number of domains. We hope that you will agree that this study is innovative and appropriate for the wide readership of your journal.

Sincerely,

Philipp Homan and Daniela Schiller

Responses to Reviewers

Reviewer #1:

Comment

Review on: Aversive smell associations shape social judgment by Homan et al.

The authors present an olfactory conditioning paradigm to study the effects of aversive smell on the behavioral and neural correlates of social judgment. In a two-staged experimental design fashion, participants first underwent aversive conditioning where social information was either paired with a smell (viewing face A) or not (viewing face B). After conditioning participants were repeatedly confronted with both faces during an fMRI study where they were asked to indicate whether to attribute a (norm violating) behavior to situational factors or the personality of the target. The authors conclude that the induction of a negative perception of a person (by aversive smell conditioning) biases the attribution style towards that person in a way that general behavior is associated with more enduring and stable characteristics of that person, but blamed less on situational factors. As an explanation the authors assume that psychological distance, induced by negative smell, might account for this effect.

Overall the manuscript is nicely written and the experimental design is well described. Nevertheless I have several remarks that may question the overall validity of the findings and interpretation of results: On a more conceptual level it was unclear to me why the authors applied functional neuroimaging to support their claims?

- **Reply:** We applied neuroimaging because we were not only interested in the behavioral effects of aversive olfactory conditioning but also their neural correlates. This opens a more holistic perspective on classical social psychology phenomena such as the fundamental attribution error, (which we here studied under the novel influence of olfactory conditioning), as it considers not only the social but also the neurobiological aspect of this phenomenon.
- **Action taken:** We now include a justification of our approach in the Introduction (p. 3).

Comment

In the introduction, the authors do not provide any hypotheses on what they would expect on the level of neural activity. In the discussion the authors mention the concept of mentalizing, but do not exemplify why and how mentalizing could be implicated during the processing of the face stimuli/vignettes. Further, the findings of VLPFC and RAG are interpreted in a reverse inference fashion and, in my opinion, do not strengthen the behavioral findings. In sum, I do not see how functional neuroimaging helps the authors with their central claim.

- **Reply:** We agree with the reviewer that the reason for using neuroimaging in this study and the related hypotheses should be mentioned earlier in the manuscript (see also previous comment).

We expected that the two competing behavioral hypotheses we proposed would be reflected by the direction of activity changes in a network of regions previously implicated in mentalizing (Moran et al., 2014) and information integration (Roy et al., 2012). Mentalizing – the ability to infer mental states of others – has been shown to positively correlate with the tendency to make trait attributions (Moran et al., 2014). We therefore expected to find activity changes during attribution when comparing CS+ to CS- in regions possibly involved in mentalizing such as the dorsomedial prefrontal cortex, posterior cingulate cortex, and angular gyrus. Regarding information integration, the computation of a single meaning from different sources of information critically depends on the ventromedial prefrontal cortex (VMPFC; Roy et al., 2012). Previous work has indicated that

activity in the VMPFC is involved in the evaluation of smell (Howard et al., 2016) and in tracking the devaluation of a smell when it is described as aversive instead of appetitive (de Aurojo et al, 2005), thereby updating the meaning by a context assessment.

We hypothesized that the distinction between the overall negative affect towards the target person informing the attribution process, and the alternative prediction that participants would take any behavior as indicative of personality traits, should be evident by the direction of the activity changes at the inference stage. The less effortful and cognitively less demanding process of attributions that do not consider the situational context should result in activity decreases in the VMPFC and mentalizing network, while an inference mode that does take the situational context into account (at least for positive behaviors) should be cognitively more demanding and reflected by an activity increase in these regions.

- **Action taken:** The introduction now includes the expected neural correlates of our hypotheses (p. 3).

Comment

Second, as the main idea of the paradigm is to judge about a (norm violating) behavior, I was wondering why the authors did not ask participants for their feelings of shame or embarrassment? There are several studies that show that social norm/etiquette violations of others induce feelings of second-hand embarrassment in observers and how mentalizing processes could contribute to the emotional experience.

- **Reply:** This is a very interesting comment. However, although our paradigm did indeed include judgments on norm violating behavior, the other half of the behavioral vignettes described in fact positive behaviors. Further, we note that the behavior was presented under situational constraints, which allowed us to test the trait attribution tendency as a function of valence. Thus, we were not interested in the reaction to the positive or negative behavior itself, but rather in the question of how a situational explanation of the respective behavior was considered in the judgment.

Comment

Further, in the present study the sequence of the situational and behavioral segment is changed throughout the experiment, thereby inducing different anticipation phases. For example, having the situation before the behavior may change the trial structure, doesn't it? I would be interested on how the authors comment on that.

- **Reply:** Thank you for raising this important point. The reason we pseudo-randomized the presentation order (behavior first/situation first) for each participant was that we wanted to rule out order effects: maybe overcoming the trait attribution tendency was a mere effect of being exposed to the situational segment at the end of the vignette, right before making the judgment. Importantly, we also additionally confirmed the absence of order effects after conducting the experiment as mentioned in the Methods: "Possible order-effects were ruled out by including the factor 'experimental-order' as additional covariate in our statistical models to control for this possible confounder" (Methods, p. 8).

Comment

Another comment: the authors begin their paper with the question of how smell associations shape the way we judge other people. Although they present two additional experiments to prove the specificity of their findings (only olfactory effect or more general and extending to the visual domain) I was wondering of how specific this finding is to social stimuli at all? Do the authors have any further evidence that smell

conditioning does not impact non-social control conditions to the same degree? How specific is the social component of their paradigm?

- **Reply:** This is another important point, thank you for raising this. Indeed, we did not include a non-social condition and thus cannot rule out that aversive smell conditioning might impact non-social judgments as well. Associating an aversive smell with a non-social object and then testing whether judgments of that object are influenced by conditioning would be an interesting extension of the current study, addressing the question whether the effects may extend to non-social targets and object-related judgements.
- **Action taken:** We included this as a possible extension of the current study in section 3.3.2 (Valence and social specificity; p. 18).

Comment

Finally, I do not see why the authors chose social closeness as the explanatory variable? This is highly speculative and unless there is no measure of perceived distance/closeness this explanation rather seems vague.

- **Reply:** We agree that the association of trait attributions with social distance is speculative in our study, as we did not directly measure social distance. We tried to acknowledge this in our manuscript by noting that our findings "could possibly reflect increased psychological distance." (p. 16). What we did directly measure, however, was the tendency to make trait attributions, which was our primary outcome measure. The link between trait attributions and distance is empirically well supported by multiple studies showing that humans tend to make more trait attributions the more psychologically, socially, and physically remote the person they judge is (Nussbaum et al., 2003; Henderson et al., 2006; Nussbaum et al., 2006; Rim et al., 2009).
- **Action taken:** We included a paragraph in the Discussion, acknowledging that the interpretation of our findings in terms of social distance is speculative (p. 19).

Comment

Minor: the morality and liking aspect of the current study seems somewhat prone to the danger of social desirability.

- **Action taken:** We agree and have added this as a possible explanation for the absence of effects in this domain of the experiment in the Discussion (p. 14).

Comment

Methods: Did the authors model all fixations in the same regressor? One could assume that different fixation phases would exert their own impact. At least, sequence of situation/disposition should be modeled separately.

- **Reply:** Given that we ruled out order effects on the behavioral level (see p. 8), we decided to model just one regressor for the behavioral part and one regressor for the situational part of the task. In addition, given the limited number of trials in our experiment, we would have ended up with too few trials to reliably estimate any additional effects of order.

Comment

The thresholding procedure is no entirely clear. The authors should rather not switch between p-levels.

- **Reply:** We can confirm that we did not use different *P*-value thresholds in our analysis but rather relied on a correction for multiple comparisons on the cluster level. What might have appeared as

different thresholds are indeed the thresholds used on the cluster-level ($P < 0.05$, corrected for family-wise error) and the voxel-level ($P < 0.001$, uncorrected). We can confirm that this is a well-established standard approach in using cluster-correction in neuroimaging (e.g., Flandin and Friston, 2016).

- **Action taken:** To clarify this issue, we added information in the Methods, explaining in more detail how we corrected for multiple comparisons (p. 11).

Comment

Also, the leave one out procedure on the second level is new to me and I do not see the benefit compared to a standard random-effects GML analysis applying FWE corrections.

- **Reply:** The leave-one-out procedure was used not as a replacement of the multiple comparisons correction (we did correct for multiple comparisons, see previous comment) but rather as an additional analysis to obtain unbiased effect size estimates for the regions where we found significant activations. The issue of biased effect size estimates in neuroimaging has generally been acknowledged by the field (Kriegeskorte et al., 2009) and refer to the fact that often significant activations in brain imaging are used in a circular manner (“double-dipping”): first, to detect the location of involved brain regions, and then again to estimate the effect sizes of those activations. The leave-one-subject-out procedure is one option to avoid this issue.

Reviewer #2:

Comment

The authors examined the effects of aversive smell associations on social judgements of faces. Two faces were repeatedly presented. One face was associated with an aversive smell for three times. Afterwards, participants completed a social judgment paradigm while brain activity was recording using fMRI.

The authors show that participants were more likely to attribute different behaviours to the person as to the situation when the face was previously paired with a smell. In contrast, liking and morality judgements of the person were not affected. Specifically, in the attribution phase, brain activity in the VMPFC and right angular gyrus was lower for the face previously associated with the smell.

The authors did an additional control experiment, in which one face was associated with a negative picture. In this experiment, no effect on trait vs. state attributions was found. The others conclude that aversive smells associated with a face / person can influence trait vs. state attributions.

The topic of the paper is interesting and the methods are sound. It is quite astonishing for me that only three pairings between a face and an aversive odour is sufficient to induce changes in later social judgments of the person. In addition, the authors used a very conservative approach for detecting fMRI activity, rendering the fMRI results more convincing. However, some of the methodological approaches used for the behavioural analyses appear to be rather complex, and it is not quite clear to me why the authors chose these approaches over simpler types of analyses. Also, the conclusion of the authors might just be one possible alternative potentially explaining their results. However, I think the paper makes an interesting scientific contribution, and I have only some comments that should be considered by the authors:

- **Reply:** We thank the reviewer for the positive evaluation of our work. We reply to each comment below.

Comment

Major comments: 1.) The authors ran an additional control group using aversive visual instead of olfactory stimuli. The authors should introduce the question of modality specificity and the control group already in the introduction. In addition, they should directly compare the results of the two groups also in a graph. Did I understand correctly that the two groups did not differ from each other with respect to the trait vs. state judgements? Please be more specific on this point and discuss it more explicitly in the paper.

- **Reply:** We agree that clarification is needed here. The reviewer is correct that the difference between the groups did not reach statistical significance with respect to trait vs. state judgments as indicated in the manuscript ($F(1, 35) = 2.21, P = 0.15$). What did differ between groups, however, was the correlation between the conditioned negative affect and subsequent trait attribution; it was positive and significant when using aversive odor, and negative (but not significant) when using aversive images. Thus, the two correlation coefficients differed significantly (Fisher's $z(34) = 2.35, P = 0.02$).
- **Action taken:** To emphasize the difference in correlations, we now include a graph combining the correlation between negative affect and trait attributions of both groups in one graph (Fig. 4f). We discuss this finding in the Discussion (p. 19) and mention the second experiment and the question of modality in the Introduction (p. 4).

Comment

2.) In the introduction, the authors state that increased cognitive load or depleted motivational resources can lead to a preferred attribution of behaviour to the person instead of context. In their results, the authors observed that also an aversive odour leads to a preferred attribution of behaviour to the person. Thus, one possible explanation is that the association with an aversive odour also increased cognitive load, possibly by inhibiting the reactivated mental representation of the aversive smell when watching the CS+ face. Alternatively, the aversive odor association could have depleted motivation resources. The authors should include and discuss these and other alternative explanation for their effect. In addition, the authors do not introduce the concept of psychological distance in the introduction. They should do so and better explain why they prefer this explanation for their effects over others.

- **Reply:** Thank you for this very interesting comment. We agree that the aspects of potential cognitive load or depleted motivational resources are indeed additional possible explanations for our findings.
- **Action taken:** We included the aforementioned potential explanations for the behavioral findings in the Discussion (p. 19.). We also mention psychological distance in the Introduction (p. 2).

Comment

3.) As already stated above, the analysis of the behavioural data appears to be quite complex. For example, the authors could have simply averaged the ratings of the different emotions (after recoding the happiness ratings) to create one index. As they used a similar strategy to calculate the regression to %trait attribution, I would strongly suggest reporting only the average index of all (and not only some) emotional ratings.

- **Reply:** We agree that the behavioral analysis appears complex at first. However, we think that the decision to present the emotions separately first is justified by the potential complexity of the emotional reaction to an aversive smell itself. This is reflected by the fact that only disgust, anger, anxiety, and sadness were highly correlated (Cronbach's $\alpha = 0.85$), while the correlation dropped considerably after inclusion of reverse-coded happiness (Cronbach's $\alpha = 0.67$). Usually, Cronbach's α is categorized by different cutoff levels, indicating distinct levels of internal consistency (Kline 2000, The handbook of psychological testing (2nd ed.). London: Routledge, p 13). A decrease in α from 0.85 to 0.67 corresponds to a change in internal consistency from "good" to "questionable" suggesting that only the 4 formerly mentioned emotions were measuring the same construct of a negative affect in this study (although obviously a larger sample size would be necessary to confirm this).
- **Action taken:** In order to make it easier for the reader to follow the analysis and results, we revised the paragraph describing the emotional reaction to the smell, including our hypothesis and describing the effects before presenting the statistical evidence (p. 13).

Comment

4.) For the fMRI analyses, the only problem that I see is that the different phases of the experiment are following directly after each other, separate only by 500ms breaks. Please discuss possible problems with respect to the slow hemodynamic response function in clearly separating these different phases. This problem is also relevant because the number of trials is quite low (32?). Please discuss this point explicitly in the paper.

- **Reply:** We note that the 500 ms period corresponds to the time that the feedback screen ("Thank you") was shown after a judgment had been made, while behavioral and situational segments were

separated by intertrial intervals of 2 - 6 seconds. However, the reviewer is right that the short intertrial interval between the different judgments might have made it difficult (i.e., decreased statistical power) to detect potential neural differences between attribution, morality, and liking. Nevertheless, we did find that the beta estimates were indeed significantly lower between the attribution phase and moral and liking evaluation, suggesting that the neural effect was specific for attributional judgments.

- **Action taken:** We included text in the Methods addressing this issue (Section 2.4.2, p. 8).

Comment

5.) Although the fMRI findings appear to be robust, the behavioural relevance if the reduced brain activity during trait vs. state judgements remains unclear. Please calculate and report potential correlations/regressions between brain activity signals and behavioural outcomes.

- **Reply:** Following the reviewer's suggestion, we calculated the correlations of the beta estimates as obtained from the leave-one-subject-out procedure (for VMPFC and RAG) and the behavioral indices, i.e., the emotional reactions to the aversive conditioning (using both the distinct emotions and the average score) and the effect on trait attributions. There was a significant positive correlation between the conditioning effect (CS+ minus CS-) on sadness and the conditioning effect (CS+ minus CS-) on the neural activity in the RAG ($r = 0.5$, $P = 0.04$). All other correlations were not significant.
- **Action taken:** We added these findings to the Results (p. 16).

Comment

Minor points: 1.) Introduction, page 2: The order of the paragraphs of the introduction is quite odd. Particularly paragraph 2 ("we predicted") appears to be misplaced. Please revise the introduction so that the background and rationale of the study becomes clearer.

- **Action taken:** We removed the phrase "we predicted" and replaced it by "We hypothesized that there might be two possible outcomes" (p. 2) and revised the Introduction as suggested by the reviewer.

Comment

2.) Page 4, Experimental design: Please better explain what exactly is the within-subject manipulation in your study. I assume that you mean that every participant viewed a CS+ face and a CS- face? Which face was associated with the odour was manipulated between subjects, correct? Please specify in the main text. Please also indicate clearly what part of the experiments were done outside or inside of the scanner. Also, indicate the duration of the different experimental phases in minutes.

- **Action taken:** We clarified that the within-subject manipulation was olfactory conditioning and revised the Experimental Design section (p. 5) following the reviewer's suggestion.

Comment

3.) Page 5, Olfactometer, end of paragraph: Please specify the range of the rating scale.

- **Action taken:** We added the range of the rating scale (p. 6).

Comment

4.) Page 6, Behavioural data analysis: sadness is missing in line 2 and line 8

- **Action taken:** We added sadness in those two sentences (Section 2.3.5, p. 7).

Comment

5.) Page 6, Behavioural task: Please specify what happened during the two consecutive runs.

- **Action taken:** We clarified that each of the consecutive runs consisted of 16 trials (p. 8).

Comment

6.) Page 13, first paragraph: Please specify the number of subjects with significant also for the VMPFC.

- **Action taken:** We clarified that 17/17 subjects showed the significant cluster in the VMPFC (p. 15).

Comment

7.) Page 24, Legend to Figure 3: Typo "Compared"

- **Action taken:** We corrected the typo (p. 27).

Highlights

- Aversive smell shapes the way we judge people long after the smell has evaporated
- Behavior is explained in terms of abstract traits instead of concrete states
- Neural activity decrease in ventromedial prefrontal cortex and right angular gyrus

Aversive smell associations shape social judgment

Philipp Homan^{1,2}, Benjamin A. Ely¹, May Yuan¹, Tobias Brosch³, John Ng⁴, Yaacov Trope⁵, Daniela Schiller^{1*}

¹Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

²Translational Research Center, University Hospital of Psychiatry, University of Bern, Switzerland

³Department of Psychology, University of Geneva, Switzerland

⁴Translational and Molecular Imaging Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

⁵Department of Psychology, New York University, New York, NY 10012, USA

***Corresponding author:** 1470 Madison Avenue, New York, NY 10029, USA.

daniela.schiller@mssm.edu

Number of pages: [32](#)

Number of Figures: 4

Number of words for Abstract: 211; main text: [6272](#)

Acknowledgments

The authors declare no competing financial interests. This work was supported in part through the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai. Funding was provided by the Klingenstein-Simons Fellowship Award in the Neurosciences and the Le Foundation to DS; and the Swiss National Science Foundation (SNF 161077) to PH. The authors thank Heather Berlin, Mauricio Delgado, and Ifat Levy for advice on this study.

Abstract

Once associating another person with an unpleasant smell, how do we perceive and judge this person from that moment on? Here, we used aversive olfactory conditioning followed by a social attribution task during functional magnetic resonance imaging to address this question. After conditioning, where one of two faces was repeatedly paired with an aversive smell, the participants reported negative affect when viewing the smell-conditioned but not the neutral face. When subsequently confronted with the smell-conditioned face (without any smell), the participants tended to judge both positive and negative behaviors as indicative of personality traits rather than related to the situation. This effect was predicted by the degree of the preceding olfactory evaluative conditioning. Whole brain analysis of stimulus by stage interaction, indicated differential activation of the ventromedial prefrontal cortex and right angular gyrus to the conditioned versus the neutral person during the attribution phase only. These results suggest that negative smell associations do not simply induce a negative perception of the target person but rather bias the attribution style toward trait attributions. The fact that this bias was evident regardless of behavior valence suggests it may reflect enhanced psychological distance. Thus, the known observation of social rejection triggered by aversive smell may be driven by a shift in social attribution style.

Key words: Attribution; olfactory conditioning; smell, ventromedial prefrontal cortex; social decision-making

1. Introduction

From bullying the “smelly kid” in class to the booming fragrance industry – the impact of smell on social life is proverbial. People with an unpleasant body odor, for instance, are often stigmatized and suffer social discrimination. How do smell associations shape the way we judge other people long after the smell has evaporated? As social beings, we constantly judge other people. We observe what they do and speculate why they do it. Upon doing so, we tend to attribute behavior primarily to the person’s personality instead of considering the context. When observing Mike’s apparent rude behavior, for example, we tend to infer that Mike must be a rude person, rather than considering the context, which may offer other explanations (Heider, 1958; Jones and Harris, 1967; Ross, 1977; Gilbert and Malone, 1995). Previous research showed that cognitive load or depleted motivational [resources as well as psychological distance](#) increase the tendency to attribute behavior to the person instead of the context (Nisbett and Ross, 1980; Trope, 1982, 1986; Trope and Cohen, 1989; Trope and Liberman, 1993; Trope and Gaunt, 2000; Gawronski, 2004), but the role of aversive smell on social attribution has not been studied before. We theorized that smell may affect social judgment by shaping attribution style.

[We used aversive olfactory conditioning to establish a relationship between an aversive smell and the target person. Subsequently, without the smell, we tested how evaluative smell conditioning influenced the process of drawing personality attributions from observed behavior, as well as liking and morality judgments. Participants were instructed to judge positive and negative behaviors that included a situational constraint during functional magnetic resonance imaging \(fMRI\), which allowed us to test the trait attribution tendency as a function of behavior valence. We hypothesized two possible outcomes. One possibility was that olfactory conditioning would result in more negative and fewer positive personality attributions for the target person, consistent with an overall negative affect towards the target person that informs the attribution process. The alternative prediction was that participants would take any behavior as indicative of personality traits, resulting in more personality attributions irrespective of behavior valence.](#)

As for the liking and morality judgments, previous research have shown that these could be influenced by aversive odors (Kirk-Smith et al., 1983; Rozin and Fallon, 1987; Alaoui-Ismaili et al., 1997;

Herz et al., 2004; Schnall et al., 2008; Stevenson, 2010; Yeshurun and Sobel, 2010; Croy et al., 2011; Reicher et al., 2016), but whether odor evaluative conditioning would similarly shape morality and liking judgments is unclear.

We combined the behavioral manipulation with neuroimaging because we were not only interested in the behavioral effects of aversive olfactory conditioning on social judgments but also their neural correlates. We expected that the two competing behavioral hypotheses we proposed would be reflected by the direction of activity changes in a network of brain regions previously implicated in mentalizing (Moran et al., 2014) and information integration (Roy et al., 2012). Mentalizing – the ability to infer mental states of others – has been shown to positively correlate with the tendency to make trait attributions (Moran et al., 2014). We therefore expected to find activity changes during attribution when comparing CS+ to CS- in regions possibly involved in mentalizing such as the dorsomedial prefrontal cortex, posterior cingulate cortex, and angular gyrus. Regarding information integration, the computation of a single meaning from different sources of information critically depends on the ventromedial prefrontal cortex (VMPFC; Roy et al., 2012). Previous work has indicated that activity in the VMPFC is involved in the evaluation of smell (Howard et al., 2016) and in tracking the devaluation of a smell when it is described as aversive instead of appetitive (de Aurojo et al., 2005), thereby updating the meaning by a context assessment.

We hypothesized that the distinction between the overall negative affect towards the target person informing the attribution process, and the alternative prediction that participants would take any behavior as indicative of personality traits, should be evident by the direction of the activity changes at the inference stage. The less effortful and cognitively less demanding process of attributions that do not consider the situational context should result in activity decreases in the VMPFC and mentalizing network, while an inference mode that does take the situational context into account (at least for positive behaviors) should be cognitively more demanding and reflected by an activity increase in these regions.

Finally, to also address the specificity of aversive olfactory conditioning, we conducted a second experiment using visual aversive stimuli to test whether the effect on social judgment specific to aversive

smell. Given that odors can induce strong affects (Kirk-Smith et al., 1983; Rozin and Fallon, 1987; Alaoui-Ismaili et al., 1997; Herz et al., 2004; Schnall et al., 2008; Stevenson, 2010; Yeshurun and Sobel, 2010), we hypothesized that the conditioning effect induced by olfactory stimuli might be more enduring and the effect on social judgments thus specific to olfactory stimuli compared to visual stimuli.

2. Materials and methods

2.1. Participants

We recruited 19 (10 female) right-handed healthy human participants without any medical conditions. One female and one male participant were excluded due to scanner problems during conditioning and the social judgment task, respectively, resulting in missing imaging data. The final sample thus consisted of 17 participants (9 female) with a mean age of 27.7 (3.9) years. All participants provided written informed consent and were financially compensated for their participation. The experiment was approved by the Internal Review Board of the Icahn School of Medicine at Mount Sinai.

2.2. Experimental Design

This was a within-subjects design in healthy volunteers involving olfactory conditioning [outside the scanner \(duration: about 12 minutes\)](#) and a social judgment paradigm [\(duration: about 16 minutes\)](#) during functional magnetic resonance imaging (fMRI) [in a single session. Smell was delivered only during the conditioning phase.](#) Participants first underwent olfactory conditioning (Fig. 1a). Before and after the conditioning, participants rated their emotions. Immediately afterwards, they completed a social judgment paradigm during fMRI (Fig. 2a). [The within-subject manipulation was olfactory conditioning, as each participant was repeatedly exposed to both the CS+ person \(paired with the smell\) and the CS- person \(never paired with the smell\) during conditioning and then again during the social judgment paradigm \(when no smell was delivered\).](#)

2.3 Olfactory conditioning

2.3.1. Stimuli

For the olfactory conditioning paradigm, two male faces with neutral facial expression were used, both of which were taken with an identical lighting source and camera angle (Extended Yale Database B).

2.3.2. Conditioning task

Two male faces (“Mike” and “Steve”) were shown in a pseudo-randomized order 9 times each (for 6 seconds). One face (conditioned stimulus; CS+) terminated with an aversive odor (unconditioned stimulus; US) in 1/3 of the trials while the other (CS-) was never paired with the odor. Between each face a fixation cross was shown for an intertrial interval varying between 6 and 10 seconds. The odor was delivered for 6 seconds per US trial. After the odor delivery a fixation cross was shown for another 14 seconds during which unscented air was delivered to the mask. The experimental order was randomized and counterbalanced across participants so that either Mike or Steve was paired with the odor. E-Prime 2.0 (Psychology Software Tools Inc., Pittsburgh, PA; <http://www.pstnet.com>) was used as presentation software.

2.3.3. Olfactometer

Participants wore a phantom nasal CPAP mask (SleepNet Corporation, Hampton, NH) and were exposed to the odor through an in-house built computerized 12-channel olfactometer based on the principles of air dilution olfactometry (Ng et al., 2011). The olfactometer consisted of a controlled valves unit, a signal control unit, a PC laptop computer, an air compressor, and a vacuum pump. The control interface programs were written in LabVIEW (a graphical programming language from National Instruments – Austin, Texas) and communicated with the olfactometer through NI USB-6221. Unscented air was delivered to the mask throughout the conditioning task unless a trial terminated with a US, in which case, the mask was filled with the aversive odor. Thus, unscented air was the neutral baseline to which the aversive odor was compared. It took approximately 400 ms to fill the mask with the odor and about the same time to vacuum it out. The odor was a commercially available flatulence odor consisting of an aqueous solution of ammonium sulfide that emitted hydrogen sulfide, a component of natural flatulence odor, when exposed to air. The odor was physically harmless but unpleasant. The unpleasantness of the smell was validated in a post-conditioning questionnaire (range: 1 – 8); participants rated the smell as intense (M=5.5, SD=1.6), aversive (M=5.3, SD=1.6), and disgusting (M=5.2, SD=2.6).

2.3.4. Behavioral Assessment

The participants rated their emotions towards Mike and Steve on a computer screen, before and after the conditioning paradigm. They were asked to rate their emotions of anger, anxiety, disgust, sadness and happiness (i.e., “emotional ratings”) on a discrete visual analogue scale between 1 and 8.

2.3.5. Behavioral data analysis

The outcome measure was the rating on the discrete visual analogue scale. The effect of olfactory conditioning was assessed by modeling the effects of emotion (5 levels; anger, anxiety, disgust, [sadness](#), happiness), stimulus (2 levels; CS+, CS-), and time (2 levels; before, after) on the outcome measure in a linear mixed model with restricted maximum likelihood estimation. A random intercept as well as slopes for emotion, stimulus and time, were included in the model to account for the intra-subject correlation. The Kenward-Roger method as provided by SAS was applied to calculate the denominator degrees of freedom. We expected to see an increase from pre- to post-conditioning in emotional ratings for the CS+ but not the CS- person for the negative emotions (anger, anxiety, disgust, [sadness](#)), and decreased ratings for happiness. Statistically, this corresponds to a significant 3-way interaction of emotion-by-stimulus-by-time.

2.4. Social judgment paradigm

2.4.1. Stimuli

For the social judgment task, we used 32 vignettes that consisted of scenarios describing the behavior of two named characters, “Mike” or “Steve”, in a given situation. The behavioral segment described a certain behavior (e.g., “Mike left the restaurant in a hurry without tipping the waitress”), whereas the situational segment described the circumstances under which the behavior took place (e.g., “Mike's baby was screaming”). One half of the vignettes described a positive behavior, the other half a negative behavior. Notably, each situational segment was constructed in a way that could potentially relativize both the positive and the negative behaviors. Behavioral and situational segments were presented separately, and the presentation order was counterbalanced across scenarios.

2.4.2. Behavioral Task

During two consecutive runs of 16 trials each, participants were asked to read brief vignettes about either Mike or Steve consisting of a behavior segment, describing a positive or negative behavior, and a situation segment, describing the circumstances surrounding the behavior. Each segment was shown for 6 seconds and separated by a fixation cross that was shown for an intertrial interval varying between 2 and 6 seconds. Participants were then presented with 3 consecutive rating screens for a maximum of 10.5 seconds where they were instructed to rate (see next section) whether the behavior was due to the person's personality (i.e., dispositional attribution) or the circumstances (i.e., situational attribution), as well as the morality of the behavior and the liking of person. Each rating screen was replaced by a feedback screen ("Thank You") for 500 ms after the judgment was made, and a final feedback screen ("Thank you. Your responses were recorded") after the last rating screen appeared for the remainder of a maximum 10.5 seconds period. Note that although the intertrial interval between the different types of social judgment (attribution, morality, and liking) was short, it appeared to be sufficient for detecting differential neural effects on social judgments (see below). During these rating and feedback screens, the face was always presented. Finally, a fixation cross was shown for an intertrial interval varying between 2 and 6 seconds. Each of the vignettes corresponded to either Mike or Steve, whose corresponding face was shown on each of the segment screens during a given vignette (Fig 2a). The experimental order was pseudo-randomized and counterbalanced across participants so that the order of screens differed between participants; that is, across all trials, there were vignettes that started with a behavioral segment and others with a situational segment, the order of the rating screens differed across trials, and the pairing of either Mike or Steve with a particular vignette varied between subjects. Possible order-effects were ruled out by including the factor 'experimental-order' as additional covariate in our statistical models to control for this possible confounder. E-Prime 2.0 (Psychology Software Tools Inc., Pittsburgh, PA; <http://www.pstnet.com>) was used as presentation software.

2.4.3. Behavioral Assessment

During the task, on each trial, the participants rated the causation of the behavior (“The behavior was caused by”), the morality of the person (“How moral is the person”), and the liking of the person (“how much do you like the person”), on a discrete visual analogue scale between 1 and 8. For attribution ratings, 1 indicated “Situation” and 8 indicated “Disposition.” For morality, 1 indicated “Unmoral” and 8 indicated “Moral.” For liking ratings, 1 indicated “Not at all” and 8 indicated “Very much” (Fig. 2a).

2.4.4. fMRI data acquisition

A 3 T Philips Gemini scanner and Philips standard head coil were used for data acquisition. Functional images were recorded in two consecutive scanning sessions. The sessions comprised 272 volumes each. A single-shot gradient echo EPI sequences (TR = 2.0 s; TE = 25 ms; FoV = 192 cm, flip angle = 75°) was used to obtain 43 oblique-axial slices with 2 mm thickness, a 1 mm inter-slice gap, and an in-plane resolution of 3 × 3 mm parallel to the anterior commissure-posterior commissure line.

2.4.5. Behavioral data analysis

The primary outcome measure was the number of trait attributions, calculated as the fraction of trials where a dispositional attribution was recorded out of the total number of trials. This dichotomization reflected the bipolar scale that considered judgments either situational or dispositional and has been used in previous studies on social judgments (Schiller et al., 2009; Brosch et al., 2013). The fraction was calculated for each subject, valence, and stimulus, and the final outcome value was calculated as the difference between CS+ and CS- for positive and negative behaviors. A linear mixed model was then used, including fixed effects of stimulus, valence, and their interaction as well as a random intercept and slopes for valence and stimulus with a variance components variance covariance matrix to account for the between subjects random variation of the effects of interest. The Kenward-Roger method as provided by SAS was applied to calculate the denominator degrees of freedom.

Secondary outcome measures included reaction times, and morality and liking ratings. For reaction times, fixed effects for time, valence, type of rating, and stimulus were included; in addition, a

random intercept for subjects and random slopes for time, rating-type, stimulus and valence; finally, a random intercept for item was included as well as random slopes for rating-type. A variance components variance-covariance matrix was used for these random effects. The Kenward-Roger method as provided by SAS was applied to calculate the denominator degrees of freedom.

Morality and liking ratings included fixed effects for time (continuous), stimulus (CS+, CS-), valence (negative scenario, positive scenario), and inference (dispositional, situational), as well as their respective interactions. A random intercept for subject and slopes for time, valence, stimulus, and inference were included as well as a random intercept for item. A variance components variance covariance matrix was used for these random effects. The Kenward-Roger method as provided by SAS was applied to calculate the denominator degrees of freedom.

To evaluate the relationship between the conditioning effect on emotional ratings and the number of trait attributions, a general linear model was calculated using emotional ratings as predictor and trait attributions as dependent variable. Therefore, the differences in trait attributions and the average of negative emotional ratings between CS+ and CS- were calculated for each subject. The significance threshold for these contrasts was set at $\alpha = 0.05$, two-tailed. SAS University Edition (SAS Institute, Cary, North Carolina, USA, <http://www.sas.com>) was used for all analyses.

2.4.6. fMRI data analysis

Functional data were analyzed with the SPM12 software package (Wellcome Trust Centre for Neuroimaging, London, UK; <http://www.fil.ion.ucl.ac.uk/spm>) and the toolboxes marsbar (Brett, 2002) (<http://marsbar.sourceforge.net>). Native-space images were first realigned, slice-time corrected, and coregistered to each subject's structural scan. Structural image preprocessing included segmentation, bias correction, and spatial normalization; these normalization parameters were also used to normalize the functional images. Finally, functional images were smoothed with a Gaussian kernel (8 mm FWHM) and resampled to $2 \times 2 \times 2$ mm voxels.

Data were modeled voxelwise with a general linear model (GLM) for each of the 17 participants. Both runs were included in one GLM as separate sessions. GLM regressors accounted for behavioral and situational segments (i.e., behavioral and situational encoding) as well as subjects' attribution, morality, and liking ratings for each stimulus separately. Durations for attribution, morality, and liking were modeled individually using the participants' reaction times, as this method provides the most sensitivity for decision-making tasks (Grinband et al., 2008). Trials where no response was recorded (mean: 4.5, SD: 3.4 per subject; 76 trials or 4.7% across the whole sample) were modeled in an additional error-regressor. Each regressor was convolved with the canonical hemodynamic response function provided by SPM12, and a high-pass filter with a cutoff period of 128 s and an autoregressive first order model correction for temporal autocorrelation was applied. Additional regressors were included as parametric modulators for each of the aforementioned regressors to account for the extinction effect using an exponential decay function (Buchel et al., 1998; Gottfried and Dolan, 2004). Finally, six regressors modeling affine head-motion parameters were included as additional covariates in all GLMs.

After computing contrast images for each subject, group analyses assessed random-effects across all 17 subjects using one-sample t-tests, testing for both positive and negative blood oxygen level dependent (BOLD) effects in the whole brain. The contrast of interest was the stimulus-effect ($CS+ > CS-$) during the attribution phase, which we compared with the stimulus effect during encoding and during liking and morality ratings. With this contrast we assessed not only the strength of the stimulus effect during attribution but also whether it was specific for the attribution phase. We report any responses that survived a whole-brain correction for familywise error at the cluster level ($P < 0.05$), with a threshold at the voxel level of $P < 0.001$, which is the standard approach for cluster-correction (Flandin and Friston, 2016).

To get an unbiased estimate of the underlying effect sizes for each significant cluster, we used a leave one subject out cross-validation procedure (Esterman et al., 2010). We repeatedly recalculated the second level GLM, iteratively leaving out one of the subjects in each calculation. The resulting GLM was then used to define regions of interest for the left out subject, thereby providing an unbiased localizer for

the extraction of the beta estimates. We used the same voxel level threshold of $P < 0.001$ as in the original analysis and defined the clusters by using the minimal Euclidian distance from the original significant clusters obtained from the full sample GLM.

3. Results

3.1. Olfactory conditioning

The participants were expected to change their ratings from pre- to post-conditioning only for the smell-conditioned (CS+) but not the neutral person (CS-). Specifically, we had hypothesized that after olfactory conditioning, participants would associate the CS+ but not the CS- person with the negative emotional reaction to the smell. Indeed, the participants reported a range of negative feelings toward the CS+ person. We therefore assessed the interaction of emotion-by-stimulus-by-time and found that it was significant ($F(4, 60.5) = 3.99, P = 0.006$; Fig. 1b). Anger and anxiety increased and happiness decreased significantly more for the conditioned compared to the neutral person ($P < 0.05$) while the increase did not quite reach statistical significance for disgust ($P < 0.07$) and was not significant for sadness ($P = 0.4$). Since the conditioning effect on anxiety, disgust, anger and sadness showed a high positive correlation (Cronbach's $\alpha = 0.85$), we calculated an affective negativity index using the average scores of these emotions for subsequent analyses. Note that the additional inclusion of reverse-coded happiness would have resulted in a considerable drop of the correlation (Cronbach's $\alpha = 0.67$), corresponding to a change in internal consistency from “good” to “questionable” (Kline, 2000), which is why we restricted the index to the highly correlated emotions.

3.2. Social judgment paradigm

3.2.1. Behavioral results: Attributions

Participants were more likely to attribute behavior to traits than to the situation when faced with the CS+ person ($b = 8.77, 95\% \text{ CI: } 0.45, 17.1; t(16) = 2.23, P = 0.04$; Fig. 2b). Importantly, there was no evidence for a valence-by-stimulus interaction ($P = 0.28$; Fig. 2c), suggesting that this attribution pattern did not significantly depend on the valence of the behavior. We further found that the enhanced tendency for trait attributions to the CS+ person (relative to CS-) was predicted by the amount of negative affect the CS+ person provoked (relative to CS-) during the preceding smell conditioning ($b = 4.16, 95\% \text{ CI: } 0.22, 8.11; t(16) = 2.25, P = 0.04$; Fig. 2d). Participants were also faster in making judgments of the CS+

person, reflected by shorter reaction times across attribution, morality and liking ratings ($b = -121.46$, 95% CI: -219.07, -23.86; $t(15.3) = -2.65$, $P = 0.02$).

3.2.2. Behavioral results: Morality and liking

Consistent with previous reports (Brosch et al., 2013), we found that liking ratings were higher for positive compared to negative behaviors (main effect of valence; $F(1, 32.4) = 41.12$, $P < 0.0001$), and this difference was more pronounced when a dispositional compared to a situational attribution had been made (attribution-by-valence interaction; $F(1, 35.3) = 11.78$, $P = 0.002$; Fig. 2e). Similarly, the morality ratings were also higher for positive compared to negative behaviors (main effect of valence; $F(1, 34) = 86.11$, $P < 0.0001$) but this difference was again more pronounced when a dispositional compared to a situational attribution had been made (attribution-by-valence interaction; $F(1, 36.9) = 11.45$, $P = 0.002$; Fig. 2f). Taken together, liking and morality ratings were consistent with the behavior valence and the type of attribution that had been made, thereby confirming the validity of the task.

Surprisingly, we did not find effects of smell conditioning on liking or morality ratings. One could have expected that participants would overall rate the conditioned person (CS+) as less likable and less moral than the neutral person (CS-), reflecting the negative emotional ratings after conditioning. A possible explanation is that the conditioning effects did not extend to the post-conditioning phase for liking and morality ratings when no smell was delivered. *Alternatively, the morality and liking ratings might have been affected by social desirability.* We also did not find effects of conditioning specific for dispositional attribution type or negative behavior valence in liking and morality ratings. As there were fewer trials for this subcategory of liking and morality ratings, the statistical power might have been insufficient to detect these more subtle differences.

3.2.3. fMRI Results

We examined the effect of conditioning (CS+ minus CS-) during the attribution phase, when participants had to decide whether the behavior was caused by dispositional or situational factors. We also examined whether this effect was specific for the attribution phase by comparing it with the stimulus

effect during encoding, when the behavioral and situational information was presented, and with the stimulus effect when liking and morality ratings were made. With this stimulus-by-stage interaction we thus tested the strength of the stimulus effect during attribution and its specificity. A whole-brain analysis of the stimulus-by-stage interaction found lower responding to the CS+ person compared to the CS- person in the VMPFC and the right angular gyrus (RAG) during the attribution stage but not during encoding ($P < 0.05$, cluster-corrected; Fig. 3a).

To visualize and quantify differential BOLD activation to the stimuli in the different stages, we obtained an unbiased effect size estimate using a leave-one-subject-out procedure and defined the VMPFC and RAG for each subject independently, and then extracted the average beta estimates for each subject (Fig. 3b). For the VMPFC, the cluster obtained remained significant at the cluster level ($P < 0.05$) in 17/17 subjects; for the RAG, the cluster was significant at the cluster level ($P < 0.05$) in 3/17 subjects and showed a trend level ($P < 0.1$) in one additional subject. We therefore extracted the parameter estimates at voxel level threshold of $P < 0.001$. The fact that the cluster was not significant in each GLM is expected and can be explained by the loss of one degree of freedom due to the left-out subject and thus the slightly reduced statistical power (Esterman et al., 2010).

To quantify and compare the stimulus difference at each stage, we then calculated a general linear mixed model with the beta estimates as outcome measure for each region and a predictor for stage. There was a significant effect of stage in the VMPFC ($F(4, 64) = 6.75$, $P = 0.0001$) and in the RAG ($F(4, 64) = 4.55$, $P = 0.003$). For the VMPFC, this effect was driven by significantly more negative beta estimates (i.e., lower responding to the CS+ person compared to the CS- person) during the attribution phase compared to all other stages: behavioral encoding ($t(64) = -4.43$, $P < 0.0001$), situational encoding ($t(64) = -4.64$, $P < 0.0001$), moral evaluation ($t(64) = -4.43$, $P < 0.0001$), ($t(64) = -4.23$, $P < 0.0001$), and liking evaluation ($t(64) = -3.38$, $P < 0.001$). For the RAG, the effect was similarly driven by significantly more negative beta estimates during attribution compared to all other stages: behavioral encoding ($t(64) = -2.49$, $P = 0.02$), situational encoding ($t(64) = -3.22$, $P = 0.002$), moral evaluation ($t(64) = -4.02$, $P = 0.0002$), and liking evaluation ($t(64) = -2.09$, $P = 0.04$).

Finally, we examined specific associations between the neural effects in the VMPFC and RAG and the behavioral indices, i.e., the emotional reactions to the aversive conditioning and the effect on trait attributions, using the beta estimates obtained from the leave-one-subject out procedure. There was a significant positive correlation between the conditioning effect (CS+ minus CS-) on sadness and the conditioning effect (CS+ minus CS-) on the neural activity in the RAG ($r = 0.5$, $P = 0.04$). All other correlations were not significant. This finding indicates that sadness in response to olfactory conditioning might have impacted neural processing in the RAG during attributional judgments, although this result should be interpreted with caution due to the correlational nature and the multiple comparisons involved in calculating the correlations.

Together, these findings indicate that the effect of olfactory conditioning was reflected by neural differentiation between the smell-conditioned and the neutral person, specifically during the attribution phase.

3.3. Validation of the results

3.3.1. Modality specificity: the effects visual stimuli on social attributions

Is the effect of aversive smell on social judgment specific to the olfaction modality or would it extend to any aversive stimuli? To test this we conducted an additional experiment using visual aversive stimuli and measuring electromyography (EMG) of the levator labii superioris. The levator labii superioris is used in lifting the upper lip and is implicated in the expression of disgust (Vrana, 1993). We recruited an additional independent sample of 20 healthy human participants (9 females; age 18-60 years). All participants provided written informed consent and were financially compensated for their participation. The experiment was approved by the Internal Review Board of the Icahn School of Medicine at Mount Sinai.

Participants underwent aversive conditioning using disgusting images from the International Affective Picture System (IAPS) database (Lang et al., 2008), from previous literature (Harris and Fiske, 2006), and the internet. These pictures were then pretested ($N = 5$) and selected according to their maximal

(15 US aversive stimuli) and minimal (16 US neutral stimuli) levator labii activation. Activation of the left levator labii was measured using EMG (MP150, Biopac Systems Inc) at a sampling rate of 200 Hz. Participants' faces were first abraded with an exfoliating scrub applied to a cotton pad. Circular surface EMG (Ag/AgCl) electrodes were then filled with a saline base conductive gel and affixed to the skin using 4 mm adhesive disks. Impedance between the skin and electrode was checked with predetermined acceptable levels of $<10 \Omega$.

During conditioning, one of two male faces was paired with an aversive image while the other face was paired with a neutral image (Fig. 4a). The subsequent social judgment task was the same as in the olfaction-experiment.

3.3.1.1. Results: visual conditioning

We found an increased levator labii EMG response ($F(3, 112) = 4.03, P = 0.009$; Fig. 4b) that was driven by a higher EMG response to the aversive images compared to the neutral images ($t(57) = 3.04, P = 0.02$, Bonferroni corrected). We also found a significant increase in negative affect against the conditioned but not the neutral person, evidenced by an emotion-by-stimulus-by-time interaction ($F(4, 233) = 3.77, P = 0.005$; Fig. 4c). Effects on anxiety, disgust, anger and sadness again showed a high positive correlation (Cronbach's $\alpha = 0.84$) and we thus calculated an affective negativity index using the average scores of these emotions for subsequent analyses.

3.3.1.2. Results: social judgment paradigm

While effects on morality and liking were comparable, as participants made ratings in accordance with their attributions (morality: $F(1, 545) = 2.16, P = 0.14$; liking: $F(1, 538) = 15.28, P = 0.0001$; Fig. 4d), we did not find evidence that the negative affect induced by the aversive images had an impact on the subsequent social judgment task: participants did not significantly distinguish between the CS+ and the CS- person ($F(1, 57) = 0.20, P = 0.7$; Fig. 4e) and there was no interaction of stimulus and valence ($F(1, 57) = 0.03, P = 0.9$; Fig. 4e) although a direct comparison of aversive stimulus type on attribution in the pooled subjects was not significant ($F(1, 35) = 2.21, P = 0.15$). Importantly, however, we did find a

significant interaction of aversive stimulus type and the prediction of trait attributions through negative affect (Fisher's $z(34) = 2.35$, $P = 0.02$) as the relationship between affect and trait attributions changed direction when using visual stimuli ($b = -2.32$, 95% CI: -6.06, 1.42; $t(18) = -1.31$, $P = 0.2$, Fig. 4f). Could these differences be explained by a difference in intensity of aversive odor compared to aversive visual stimuli? Although we cannot rule out this possibility, the overall comparable effects of olfactory and visual conditioning on emotional ratings speak against this explanation.

3.3.2. Valence *and* social specificity

Is the effect on social judgment specific to aversive odor or would any odor, even positive, have an effect? As our study did not include a positive or neutral control odor, we cannot answer this question. Future studies should thus compare aversive odor to neutral or positive odors when assessing the effect of smell associations on social attributions. Speculatively, given that aversive smell had an overall effect on trait attributions regardless of behavior valence, suggesting that the effect may have been mediated by enhanced psychological distance (see discussion); it is possible the positive smell may decrease psychological distance, resulting in an opposite effect of bias toward situational attributions.

As for social specificity, we did not include a non-social condition and thus cannot rule out that aversive smell conditioning might impact non-social judgments as well. Associating an aversive smell with a non-social object and then testing whether judgments of that object are influenced by conditioning would be an interesting extension of the current study, addressing the question whether the effects may extend to non-social targets and object-related judgements.

Discussion

This study shows that a negative affect against a social target induced by aversive olfactory conditioning predicted a switch to a biased attribution style: the situation was more often neglected and both positive and negative behaviors of the smell-associated person were indicative of personality traits.

When judging the CS+ person, participants simply took the behavior they observed as indication of a corresponding trait. They did not differentiate between positive or negative behaviors. This

cognitively less demanding attribution style has been shown to occur under less optimal conditions (Trope and Liberman, 1993), and could explain an overall increase in trait attributions regardless of behavior valence. The competing hypothesis of an attribution process that is mainly informed by the negative affect is thus not supported by our findings. Such an influence should have caused a stimulus-by-valence interaction: more trait attributions for negative behaviors and more context attributions for positive behaviors.

The fact that aversive olfactory conditioning induced more trait attributions irrespective of the valence of behavior could possibly reflect increased psychological distance (Nussbaum et al., 2003). Previous research found that we tend to make more trait attributions for others' compared to own behavior (Jones, 1976; Fiedler et al., 1995; Robins et al., 1996), even more so when the other person is unfamiliar (Idson and Mischel, 2001). We note, however, that the association of trait attributions with social distance is speculative in our study, as we did not measure social distance directly. Nevertheless, the link between trait attributions and distance is empirically well supported by multiple studies showing that humans tend to make more trait attributions for people who are more psychologically, socially, and physically remote (Nussbaum et al., 2003; Henderson et al., 2006; Nussbaum et al., 2006; Rim et al., 2009).

Psychological distance may also explain the specificity of the effect to the olfactory but not visual domain. Although we could successfully associate a social target with aversive visual stimuli, only the smell-induced negative affect predicted subsequent attributions. It is possible that olfaction-induced affect triggers avoidance and mental distancing (Nussbaum et al., 2003) and therefore promotes a disengaged attribution style. Future studies could provide more evidence for a possible domain-specific role in how affect modulates the perceived diagnosticity of behavior.

Alternatively, the association with an aversive odor might have increased cognitive load. For instance, whenever participants watched a CS+ face, it might have taken cognitive effort to actively inhibit the mental representation of the aversive smell that was activated by the CS+ face. The existence of mental representations of odors without specific expertise in olfaction, however, has been questioned in the past (Royet, 2013). Alternatively, the aversive odor association may have depleted motivation in

general, resulting in a higher tendency to make trait attributions, irrespective of behavior valence (Nisbett and Ross, 1980; Trope, 1982, 1986; Trope and Cohen, 1989; Trope and Liberman, 1993; Trope and Gaunt, 2000; Gawronski, 2004).

Our neural findings indicate that regions implicated in mentalizing (Vollm et al., 2006; Mason and Just, 2011; Sperduti et al., 2011; Moran et al., 2014) and information integration (Roy et al., 2012) mediated the cognitively disengaged attribution-style. When judging the conditioned person, compared to the neutral person, there was decreased activity in the RAG and the VMPFC specifically during the attribution phase. This is in line with the suggestion of two stages underlying the attribution of behavior (Trope, 1986; Gawronski and Creighton, 2013): at the identification stage, prior (assumptions about the person, e.g., Mike has a bad character), behavioral (the person's behavior, e.g., Mike left the restaurant without tipping the waitress) and situational (the context, e.g., Mike's baby was screaming) cues are categorized in trait-relevant terms (e.g., rudeness); at the attribution stage, these terms guide the inference-process (e.g., Mike is rude) with prior and behavior information having an additive and situational information having a subtractive effect. Importantly, identification is considered unintentional and effortless while the actual attribution involves an effortful recomputation of the encoded information (Trope, 1986; Gawronski, 2004). Supporting this distinction, the neural activity decrease was evident at the attribution stage but not during encoding.

Our finding of decreased VMPFC activity fits nicely with this region's role in contextualization. Activity in the VMPFC is involved in the evaluation of smell (Howard et al., 2016) and in tracking the devaluation of a smell when it is described as aversive instead of appetitive (de Araujo et al., 2005), thereby updating the meaning by a context assessment. Lesion studies in animals and humans further support the dependence of such contextual updates on the VMPFC (Fellows and Farah, 2003; Noonan et al., 2010). The emerging "meaning-centered" perspective (Roy et al., 2012) has thus suggested that the VMPFC is a crucial hub in contextualizing a perception, including perceptions related to the self and others (Harris et al., 2005; Gilbert et al., 2006; Harris and Fiske, 2007; Denny et al., 2012). From a social cognition perspective, the VMPFC has been implicated in the representation and evaluation of others

perceived to be similar to the self (Delgado et al., 2016). Specifically, activation of the VMPFC was found during evaluations of other people when they were emotionally close or shared traits with the self (Mitchell et al., 2006; Murray et al., 2012).

To conclude, this study showed that an aversive smell may be used to associate a negative affect with one but not the other of two formerly neutral and unknown persons, thereby creating an orthogonal negative affect for a particular identity. Most importantly, this does not lead to a corresponding negative assessment of this person. Instead, even positive behavior becomes diagnostic for the personality, depending on the strength of the negative affect and specifically for aversive smell. The affective influence on social judgment thus goes beyond valence: promoted by a neural activity decrease in the RAG and the VMPFC, the result is a switch to a disengaged attribution style. Consistent with the known observation of social rejection due to smell, our study shows that smell associations do indeed affect social judgment in a way that could be seen as social distancing.

References

- Alaoui-Ismaili O, Robin O, Rada H, Dittmar A, Vernet-Maury E (1997) Basic emotions evoked by odorants: comparison between autonomic responses and self-evaluation. *Physiol Behav* 62:713-720.
- Brett M, Anton, J.L., Valabregue, R., Poline, J.B. (2002) Region of interest analysis using an SPM toolbox. In: 8th International Conference on Functional Mapping of the Human Brain. Sendai, Japan.
- Brosch T, Schiller D, Mojdehbakhsh R, Uleman JS, Phelps EA (2013) Neural mechanisms underlying the integration of situational information into attribution outcomes. *Soc Cogn Affect Neurosci* 8:640-646.
- Buchel C, Morris J, Dolan RJ, Friston KJ (1998) Brain systems mediating aversive conditioning: an event-related fMRI study. *Neuron* 20:947-957.
- Croy I, Olgun S, Joraschky P (2011) Basic emotions elicited by odors and pictures. *Emotion* 11:1331-1335.
- de Araujo IE, Rolls ET, Velazco MI, Margot C, Cayeux I (2005) Cognitive modulation of olfactory processing. *Neuron* 46:671-679.
- Delgado MR, Beer JS, Fellows LK, Huettel SA, Platt ML, Quirk GJ, Schiller D (2016) Viewpoints: Dialogues on the functional role of the ventromedial prefrontal cortex. *Nat Neurosci* 19:1545-1552.
- Esterman M, Tamber-Rosenau BJ, Chiu YC, Yantis S (2010) Avoiding non-independence in fMRI data analysis: leave one subject out. *Neuroimage* 50:572-576.
- Flandin G, Friston K. Analysis of family-wise error rates in statistical parametric mapping using random field theory. [arXiv:1606.08199 \[stat.AP\]](https://arxiv.org/abs/1606.08199).
- Gawronski B (2004) Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology* 15:183-217.

Gawronski B, Creighton LA (2013) Dual Process Theories. In: The Oxford handbook of social cognition (Carlston DE, ed), pp 282-312. New York, NY: Oxford University Press.

Gilbert DT, Malone PS (1995) The Correspondence Bias. *Psychol Bull* 117:21-38.

Gottfried JA, Dolan RJ (2004) Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nat Neurosci* 7:1144-1152.

Grinband J, Wager TD, Lindquist M, Ferrera VP, Hirsch J (2008) Detection of time-varying signals in event-related fMRI designs. *Neuroimage* 43:509-520.

Harris LT, Fiske ST (2006) Dehumanizing the lowest of the low: neuroimaging responses to extreme out-groups. *Psychol Sci* 17:847-853.

Heider F (1958) *The Psychology of Interpersonal Relations*. New York, NY: Wiley.

Herz RS, Schankler C, Beland S (2004) Olfaction, Emotion and Associative Learning: Effects on Motivated Behavior. *Motivation and Emotion* 28:363-383.

Howard JD, Kahnt T, Gottfried JA (2016) Converging prefrontal pathways support associative and perceptual features of conditioned stimuli. *Nat Commun* 7:11546.

Jones EE, Harris VA (1967) The attribution of attitudes. *Journal of experimental social psychology* 3:1-24.

Kirk-Smith MD, Van Toller C, Dodd GH (1983) Unconscious odour conditioning in human subjects. *Biol Psychol* 17:221-231.

[Kline, P \(2000\) Handbook of psychological testing. London: Routledge.](#)

Lang PJ, Bradley MM, Cuthbert BN (2008) International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. In. Gainesville, FL: University of Florida.

Mason RA, Just MA (2011) Differentiable cortical networks for inferences concerning people's intentions versus physical causality. *Hum Brain Mapp* 32:313-329.

Mitchell JP, Macrae CN, Banaji MR (2006) Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50:655-663.

- Moran JM, Jolly E, Mitchell JP (2014) Spontaneous mentalizing predicts the fundamental attribution error. *J Cogn Neurosci* 26:569-576.
- Murray RJ, Schaer M, Debbane M (2012) Degrees of separation: a quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self- and other-reflection. *Neurosci Biobehav Rev* 36:1043-1059.
- Ng J, Evaes E, Carpenter D, Tang C (2011) An MR compatible olfactometer for clinical research use. In: *Proc. Intl. Soc. Mag. Reson. Med.*, p 3659.
- Nisbett RE, Ross L (1980) *Human inference: strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.
- Nussbaum S, Trope Y, Liberman N (2003) Creeping dispositionism: the temporal dynamics of behavior prediction. *J Pers Soc Psychol* 84:485-497.
- Reicher SD, Templeton A, Neville F, Ferrari L, Drury J (2016) Core disgust is attenuated by ingroup relations. *Proc Natl Acad Sci U S A* 113:2631-2635.
- Ross L (1977) The intuitive psychologist and his shortcomings: Distortions in the attribution process. In: *Advances in social psychology* (Berkowitz L, ed), pp 173-220. New York, NY: Academic Press.
- Roy M, Shohamy D, Wager TD (2012) Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn Sci* 16:147-156.
- Royet JP, Plailly J, Saive AL, Veyrac A, Delon-Martin C (2013) [The impact of expertise in olfaction. *Front Psychol* 4:928.](#)
- Rozin P, Fallon AE (1987) A perspective on disgust. *Psychological Review* 94:23-41.
- Schiller D, Freeman JB, Mitchell JP, Uleman JS, Phelps EA (2009) A neural mechanism of first impressions. *Nat Neurosci* 12:508-514.
- Schnall S, Haidt J, Clore GL, Jordan AH (2008) Disgust as embodied moral judgment. *Personality & social psychology bulletin* 34:1096-1109.
- Sperduti M, Delaveau P, Fossati P, Nadel J (2011) Different brain structures related to self- and external-agency attribution: a brief review and meta-analysis. *Brain structure & function* 216:151-157.

- Stevenson RJ (2010) An Initial Evaluation of the Functions of Human Olfaction. *Chemical Senses* 35:3-20.
- Trope Y (1982) Inferences of personal characteristics on the basis of information retrieved from one's memory: Cambridge University Press.
- Trope Y (1986) Identification and Inferential Processes in Dispositional Attribution. *Psychological Review* 93:239-257.
- Trope Y, Cohen O (1989) Perceptual and inferential determinants of behavior-correspondent attributions. *Journal of experimental social psychology* 25:142-158.
- Trope Y, Liberman N (1993) The use of trait conceptions to identify other people's behavior and to draw inferences about their personalities. *Personality & social psychology bulletin* 19:553-562.
- Trope Y, Gaunt R (2000) Processing Alternative Explanations of Behavior: Correction or Integration? *Journal of Personality and Social Psychology* 79:344-354.
- Vollm BA, Taylor AN, Richardson P, Corcoran R, Stirling J, McKie S, Deakin JF, Elliott R (2006) Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *Neuroimage* 29:90-98.
- Vrana SR (1993) The psychophysiology of disgust: differentiating negative emotional contexts with facial EMG. *Psychophysiology* 30:279-286.
- Yeshurun Y, Sobel N (2010) An odor is not worth a thousand words: from multidimensional odors to unidimensional odor objects. *Annu Rev Psychol* 61:219-241, C211-215.

Figure Legends

Figure 1

(a) Task design of aversive olfactory conditioning. During 18 trials, two male faces were shown in a pseudo-randomized order for 6 s, followed by a fixation cross (intertrial interval 6-10 s). One of the faces (selection counterbalanced) terminated with an aversive odor in one-third of the trials. The odor was delivered for 6 seconds per US trial. After the odor delivery a fixation cross was shown for another 14 seconds during which unscented air was delivered to the mask. (b) Evaluative conditioning. Before and after the conditioning, emotional ratings were collected. After conditioning, significant increases were evident in anger and anxiety and a significant decrease in happiness. Error bars indicate standard errors of the mean.

Figure 2

(a) Task design of social attributions. There were 32 vignettes consisting of a behavioral and a situational segment (6 s each; intertrial interval 2 - 6 s; order counterbalanced). Following each vignette, the participants answered questions about the cause of the behavior and the morality and liking of the person (maximum duration 10.5 s; questions order counterbalanced). One half of the vignettes described a positive behavior, the other half a negative behavior. Notably, each situational segment was constructed in a way that could potentially relativize both the positive and the negative behaviors. (b) Irrespective of behavior valence, participants were more likely to attribute behavior to personality traits when they had to explain the behavior of the CS+ compared to the CS-. (c) No significant interaction of valence-by-stimulus was evident but a significant main effect of stimulus. Trait attributions were calculated as the fraction of trait-attribution trials divided by the number of trials. Neg, negative behavior; Pos, positive behavior. Error bars indicate standard errors of the mean. (d) The conditioning effect (CS+ minus CS-) on negative emotions (calculated as the average of disgust, anger, anxiety, and sadness ratings) towards the CS+ person predicted the amount of trait attributions. Liking (e) and morality (f) ratings, categorized by dispositional and situational attributions and negative and positive scenarios show that participants overall

rated negative compared to positive behavior as less likable and less moral, but that this difference was greater when a dispositional compared to a situational attribution had been made. Thus, participants adjusted their ratings in accordance with the type of attribution they had made, confirming the validity of the task. Error bars indicate standard errors of the mean.

Figure 3

(a) A whole-brain analysis of the contrast CS+ > CS- tested the stimulus-by-stage interaction for attribution vs. encoding found lower responding to the CS+ person compared to the CS- person in the RAG (peak activation at x, y, z = 52, -60, 24; cluster size 71 voxels) and in the VMPFC (peak activation at x, y, z = -2, 62, -4; cluster size 155 voxels) during the attribution stage but not during encoding ($P < 0.05$, cluster-corrected for family-wise error). (b) Unbiased beta estimates (using an iterative leave-one-subject-out procedure) extracted from the RAG and VMPFC for the contrast CS+ > CS- for each stage of the social judgment task. For both regions, lower responding to the CS+ person compared to the CS- person was specific for the attribution stage compared to all other stages. RAG, right angular gyrus; VMPFC, ventromedial prefrontal cortex; behav, behavior encoding; sit, situation encoding; attrib, attribution; moral, moral evaluation; like, liking evaluation. Error bars indicate standard errors of the mean.

Figure 4

Experiment 2: Aversive visual conditioning. (a) Task design. During 30 trials, two male faces were shown in a pseudo-randomized order for 4 s, followed by a fixation cross (intertrial interval 8-12 s). One of the faces (selection counterbalanced) terminated with an aversive image (US aversive), the other terminated with a neutral image (US neutral). The US images were displayed for 6 s. (b) Electromyography (EMG) of the levator labii during aversive visual conditioning. The EMG response (indexed as percentage above baseline) to the aversive images was higher compared to the neutral images. Error bars indicate standard errors of the mean. (c) Manipulation check. Before and after the conditioning, emotional ratings were collected. After conditioning, disgust and anxiety were significantly higher for the conditioned person while no significant increase was evident for anger and sadness. Error bars indicate

standard errors of the mean. For morality and liking (**d**), categorized by dispositional and situational attributions and negative and positive scenarios, we found comparable effects using aversive images compared to aversive smell. Error bars indicate standard errors of the mean. (**e**) In contrast to the olfactory experiment, we did not find evidence that the negative affect induced by the aversive images had an impact on the social judgment task, where participants did not significantly distinguish between the conditioned and the neutral person and there was no interaction of person and behavior. Error bars indicate standard errors of the mean. (**f**) The relationship between affect and trait attributions changed direction when using aversive images [instead of olfactory stimuli, resulting in a negative relationship between negative emotions and trait attributions induced by aversive images compared to a significantly more positive relationship when using aversive olfactory stimuli.](#)

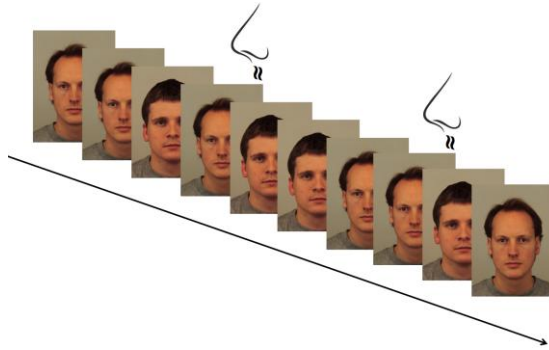
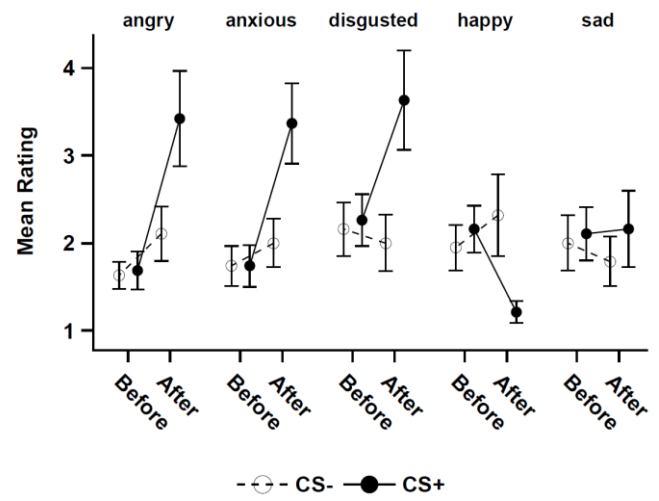
Figure 1**a****B**

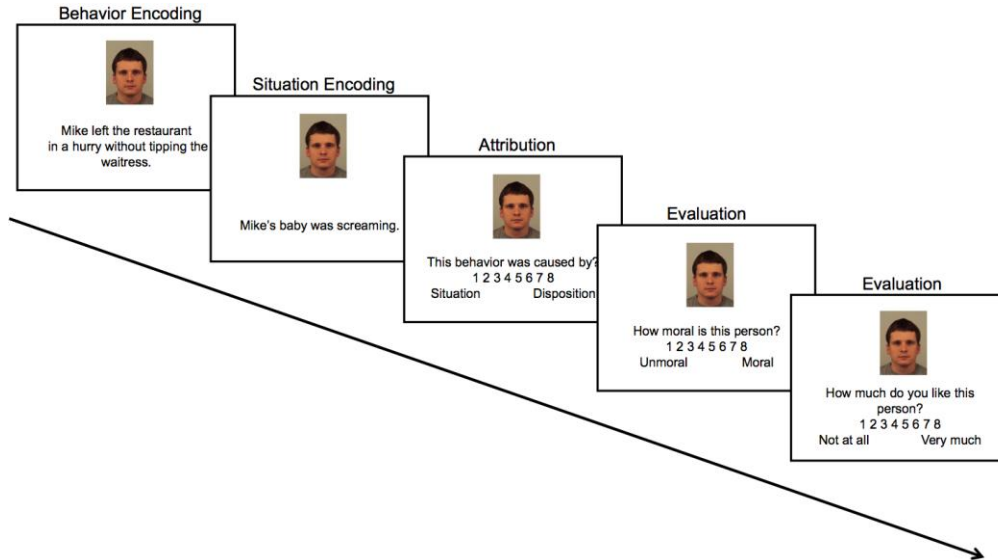
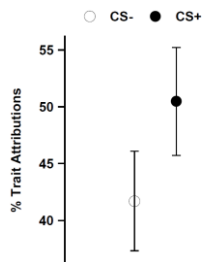
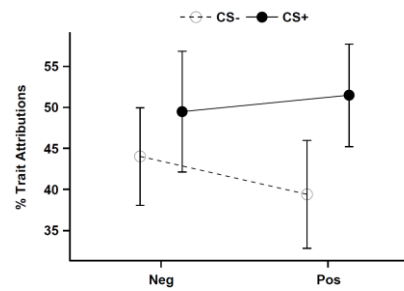
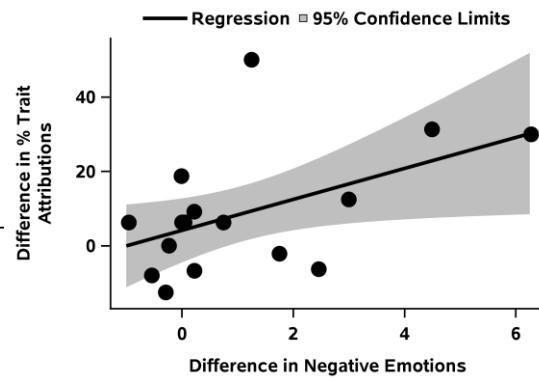
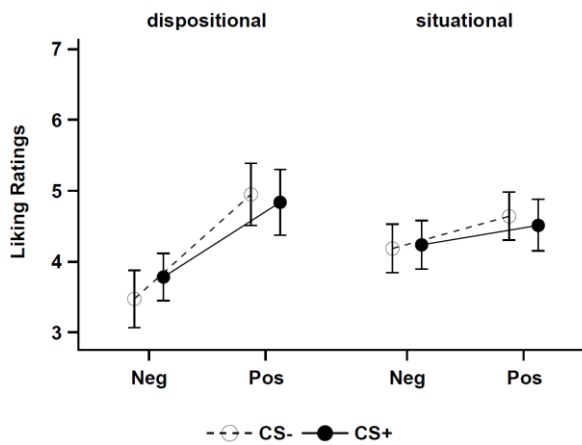
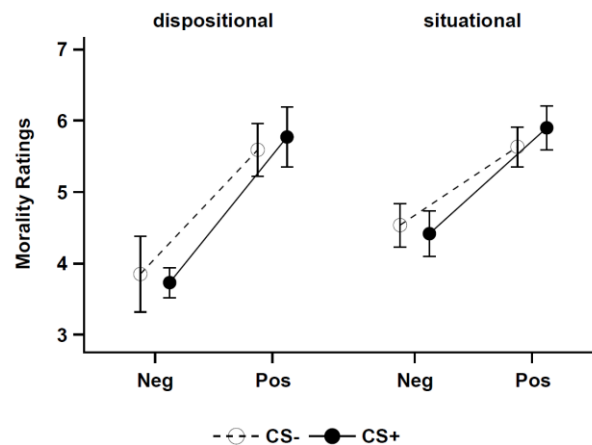
Figure 2**a****b****c****d****e****f**

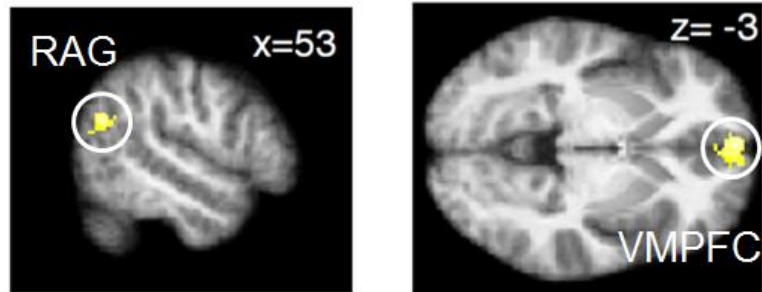
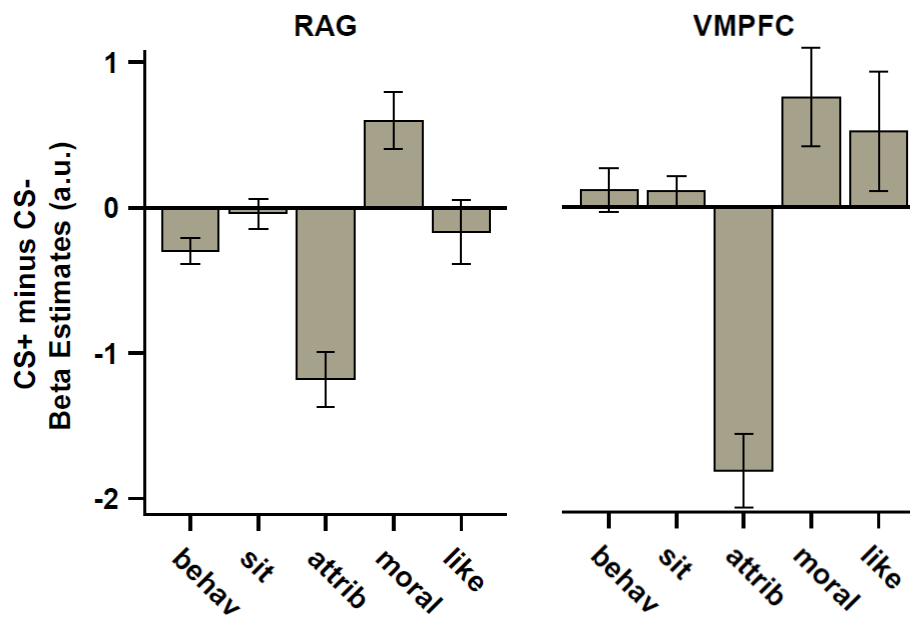
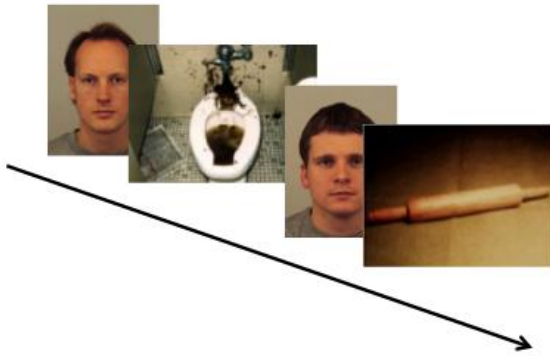
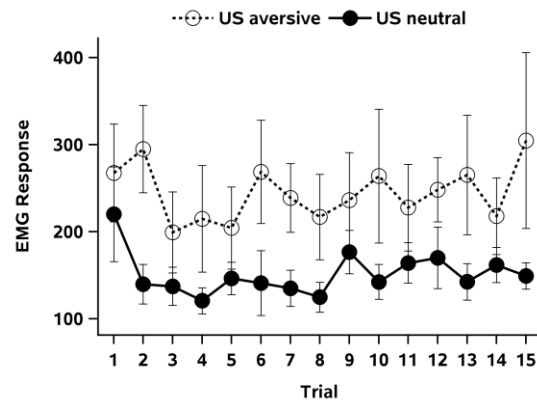
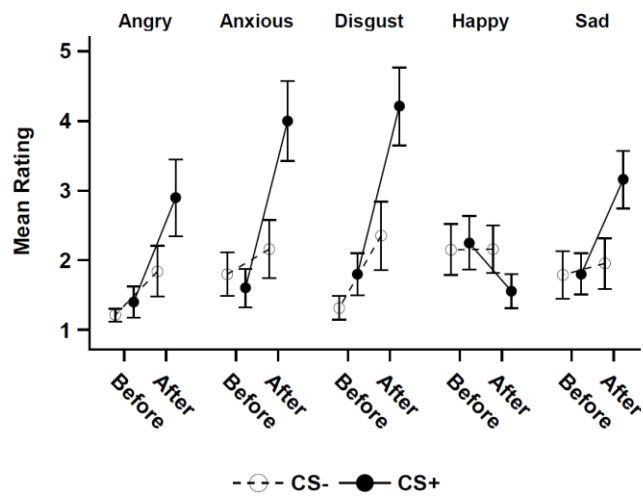
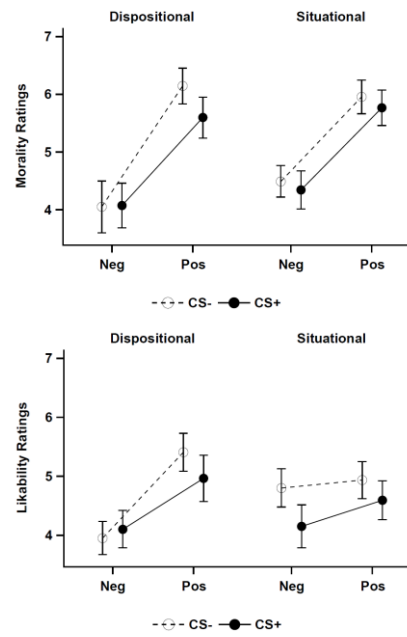
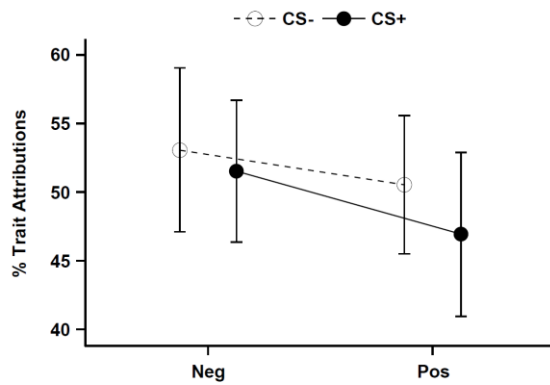
Figure 3**a****b**

Figure 4**a****b****c****d****e****f**