Elsevier Editorial System(tm) for
Neurobiology of Learning and Memory
Manuscript Draft

Manuscript Number:

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, Ph.D.

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.

Suggested Reviewers: Lasana T Harris Ph.D.
UCL London
lasana.harris@ucl.ac.uk

Thalia P Wheatley Ph.D.
Dartmouth College
thalia.p.wheatley@dartmouth.edu

Alex Todorov Ph.D.
Princeton University
atodorov@princeton.edu

Rachel Herz Ph.D.
Brown University
Rachel_Herz@brown.edu

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

April 12th, 2017

Dear Editor,

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

From bullying the "smelly kid" in class to the booming fragrance industry – the impact of smell on social life is proverbial. Here we show that an aversive smell extends the immediate gut reaction and shapes the way we judge people long after the smell has evaporated.

To measure the impact of smell associations on social judgment, we designed a task where participants had to judge two persons' positive and negative behaviors under specific situations. One of the faces had been paired with a foul smell by olfactory conditioning before the task. The aversive smell association, we predicted, should result in a higher tendency to discard the situation, and explain the behavior in terms of abstract traits instead of concrete states.

Indeed, by deactivating the ventromedial prefrontal cortex together with the right angular gyrus, participants relied primarily on the behavior when faced with the smell-associated person. This bias was evident regardless of behavior valence, suggesting that 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.

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

*Manuscript
Click here to view linked References

Aversive smell associations shape social judgment

Philipp Homan[1,2], Benjamin A. Ely[1], May Yuan[1], Tobias Brosch[3], John Ng[4], Yaacov Trope[5], Daniela Schiller[1*]

[1]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[2]Translational Research Center, University Hospital of Psychiatry, University of Bern, Switzerland
[3]Department of Psychology, University of Geneva, Switzerland
[4]Translational and Molecular Imaging Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[5]Department 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: 29

Number of Figures: 4

Number of words for Abstract: 211; main text: 4797

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 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 predicted 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.

To address these questions, 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.

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 and a social judgment paradigm during functional magnetic resonance imaging (fMRI) on a single session. 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).

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; 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, 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), 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, 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 reminder of a maximum 10.5 seconds period. 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 x 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 x 2 x 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 at the cluster level ($P < 0.05$), with a threshold at the voxel level of $P < 0.001$.

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-). 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 stimulus-by-time interaction in 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.

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% 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% 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. 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 ventromedial prefrontal cortex (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 in each GLM remained significant at the cluster level ($P < 0.05$); 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 ($t(64) = -2.09$, $P = 0.04$).

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 Ω.

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 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.

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). Thus, the more psychologically remote the person is, the less we cognitively engage in the attribution process (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.

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 right angular gyrus 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 right angular gyrus 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.

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.

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.

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 comapred 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.

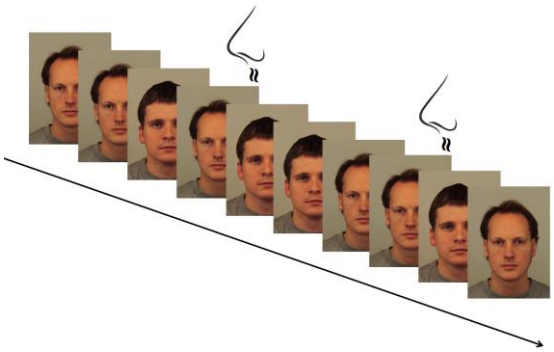
Figure 4

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, resulting in a significant interaction of aversive stimulus type and the prediction of trait attributions through negative affect.

Figure 1

a



b

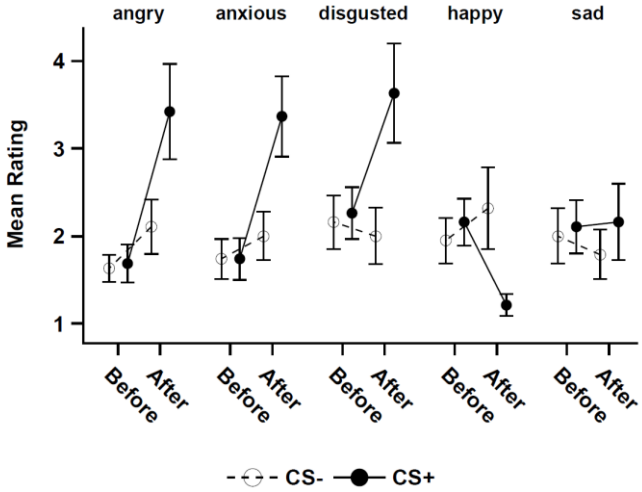

angry
anxious
disgusted
happy
sad
Mean Rating
4
3
2
1
Before
After
CS-
CS+

Figure 2

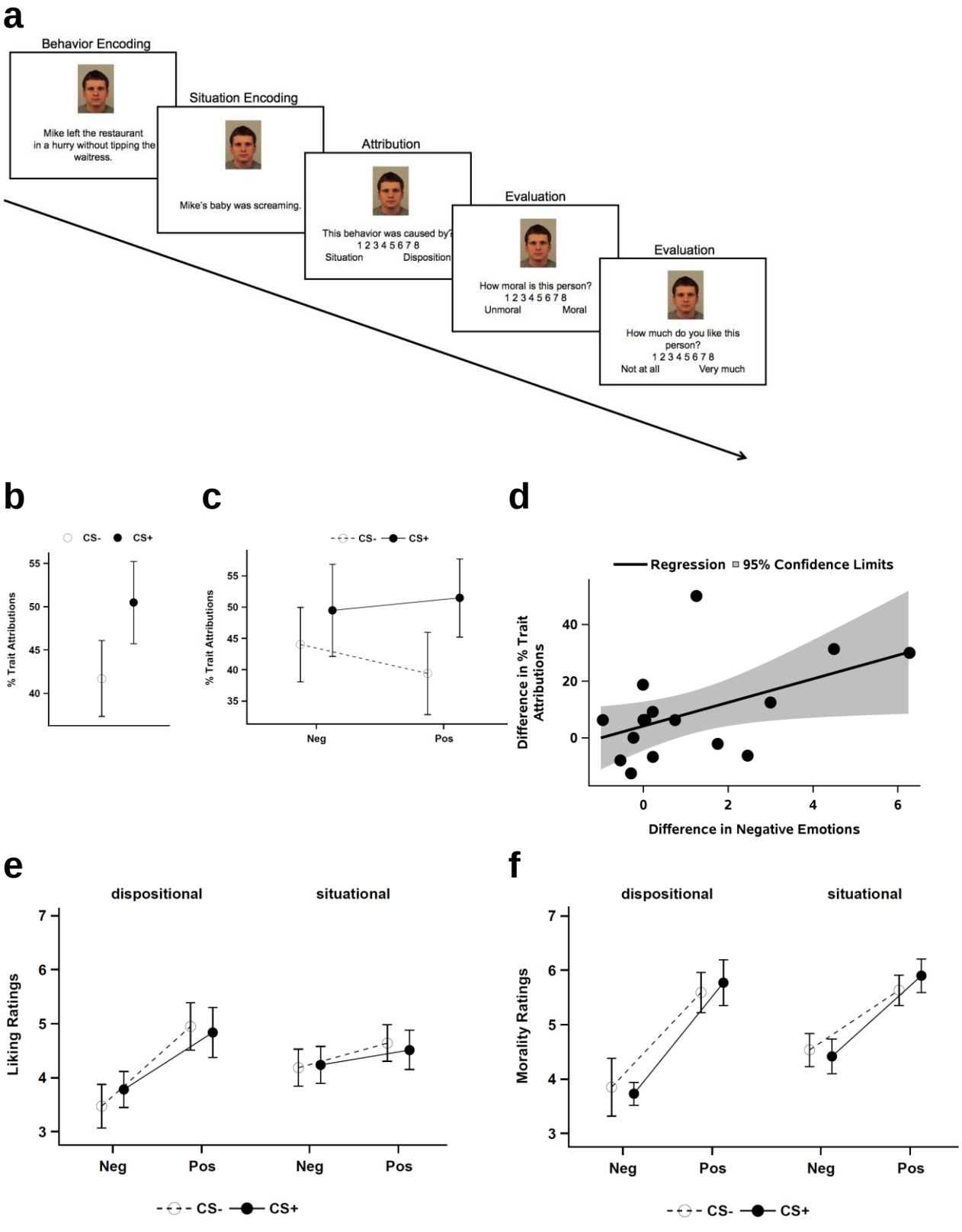
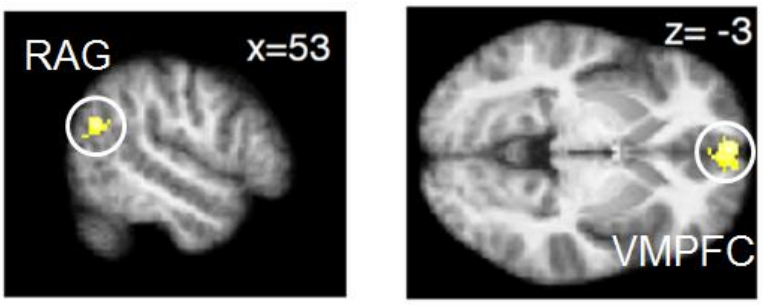

a
Behavior Encoding
Mike left the restaurant in a hurry without tipping the waitress.
Situation Encoding
Mike's baby was screaming.
Attribution
This behavior was caused by?
1 2 3 4 5 6 7 8
Situation Disposition
Evaluation
How moral is this person?
1 2 3 4 5 6 7 8
Unmoral Moral
Evaluation
How much do you like this person?
1 2 3 4 5 6 7 8
Not at all Very much
b
CS- CS+
% Trait Attributions
c
CS- CS+
% Trait Attributions
Neg Pos
d
Regression 95% Confidence Limits
Difference in % Trait Attributions
Difference in Negative Emotions
e
dispositional situational
Liking Ratings
Neg Pos Neg Pos
CS- CS+
f
dispositional situational
Morality Ratings
Neg Pos Neg Pos
CS- CS+

Figure 3

a


RAG
x=53
z= -3
VMPFC

b

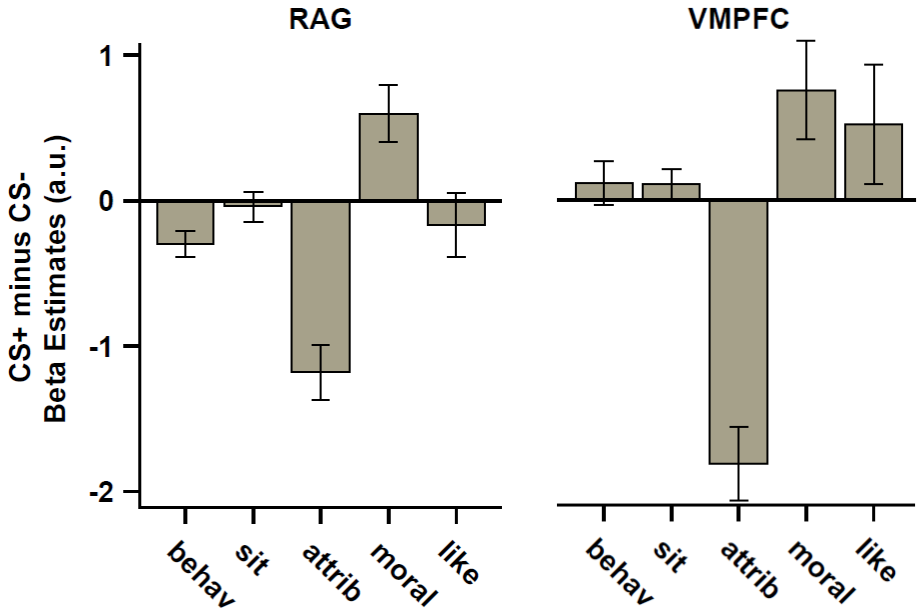
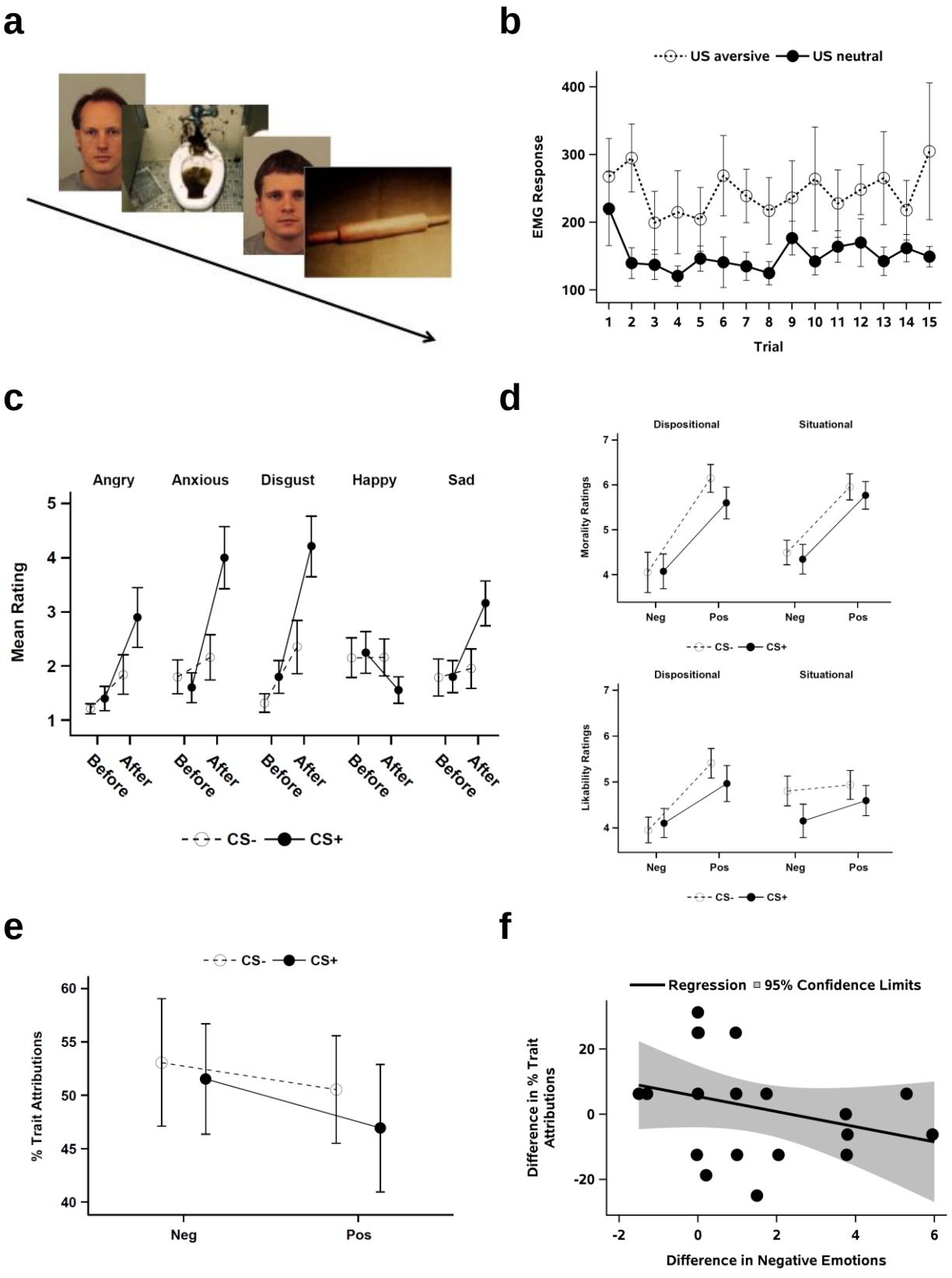

RAG
VMPFC
CS+ minus CS-
Beta Estimates (a.u.)
1
0
-1
-2
behav
sit
attrib
moral
like

Figure 4


a
b
US aversive
US neutral
EMG Response
Trial
c
Angry
Anxious
Disgust
Happy
Sad
Mean Rating
Before
After
CS-
CS+
d
Dispositional
Situational
Morality Ratings
Likability Ratings
Neg
Pos
e
% Trait Attributions
f
Regression
95% Confidence Limits
Difference in % Trait Attributions
Difference in Negative Emotions