



**HAL**  
open science

## Discriminating between sick and healthy faces based on early sickness cues: an exploratory analysis of sex differences

Arnaud Tognetti, Evelina Thunell, Marta Zakrzewska, Jonas Olofsson, Mats Lekander, John Axelsson, Mats J Olsson

► **To cite this version:**

Arnaud Tognetti, Evelina Thunell, Marta Zakrzewska, Jonas Olofsson, Mats Lekander, et al.. Discriminating between sick and healthy faces based on early sickness cues: an exploratory analysis of sex differences. *Evolution, Medicine, and Public Health*, In press, 10.1093/emph/eoad032/7284093 . hal-04235692v1

**HAL Id: hal-04235692**

**<https://hal.inrae.fr/hal-04235692v1>**

Submitted on 10 Oct 2023 (v1), last revised 7 Nov 2023 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

**Discriminating between sick and healthy faces based on early sickness cues: an exploratory analysis of sex differences**

Arnaud Tognetti<sup>1,2</sup>, Evelina Thunell<sup>1</sup>, Marta Zakrzewska<sup>1</sup>, Jonas Olofsson<sup>3</sup>, Mats Lekander<sup>1,4,5</sup>,  
John Axelsson<sup>1,4</sup>, and Mats J. Olsson<sup>1</sup>

<sup>1</sup>Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

<sup>2</sup>CEE-M, CNRS, INRAE, Institut Agro, University of Montpellier, Montpellier, France

<sup>3</sup>Department of Psychology, Stockholm University, Stockholm, Sweden

<sup>4</sup>Stress Research Institute, Department of Psychology, Stockholm University, Stockholm, Sweden

<sup>5</sup>Osher Center for Integrative Health, Karolinska Institutet, Stockholm, Sweden

*Corresponding author:*

Arnaud Tognetti  
Division of Psychology  
Department of Clinical Neuroscience  
Karolinska Institutet  
Nobels väg 9, 171 77  
Stockholm, Sweden  
Email: arnaud.tognetti@ki.se

© The Author(s) 2023. Published by Oxford University Press on behalf of the Foundation for Evolution, Medicine, and Public Health.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## ABSTRACT

### Background and objectives:

It has been argued that sex and disease-related traits should influence how observers respond to sensory sickness cues. In fact, there is evidence that humans can detect sensory cues related to infection in others, but lack of power from earlier studies prevents any firm conclusion regarding whether perception of sickness cues is associated with sex and personality traits. Here, we tested whether women (relative to men), individuals with poorer self-reported health, and who are more sensitive to disgust, vulnerable to disease, and concerned about their health, overestimate the presence of, and/or are better at detecting sickness cues.

### Methodology:

In a large online study, 343 women and 340 men were instructed to identify the sick faces from a series of sick and healthy photographs of volunteers with an induced acute experimental inflammation. Participants also completed several disease-related questionnaires.

### Results:

While both men and women could discriminate between sick and healthy individuals above chance level, exploratory analyses revealed that women outperformed men in accuracy and speed of discrimination. Further, we demonstrated that higher disgust sensitivity to body odors is associated with a more liberal decision criterion for categorizing faces as sick.

### Conclusion:

Our findings give strong support for the human ability to discriminate between sick and healthy individuals based on early facial cues of sickness and suggest that women are significantly, although only slightly, better at this task. If this finding is replicated, future studies should determine whether women's better performance is related to increased avoidance of sick individuals.

### Lay summary

Researchers suggest that humans evolved the ability to detect sickness in others, which aids their avoidance of sick individuals, and thereby reduces contagion risk. Our study shows that humans distinguish healthy from sick individuals by looking at their face, and women are faster and slightly better at it than men.

## 1. Introduction

While human mortality rates from infectious diseases have decreased globally over time, infectious disease remains a major source of mortality worldwide. In fact, fatal epidemics (e.g. influenza, ebola, SARS, and COVID-19) are expected to increase in occurrence with climate change and globalization [1]. Thus, determining the behavioral mechanisms that prevent interpersonal contagion could guide public health interventions focused on infectious-disease transmission.

Due to their long coexistence with pathogens, humans and other extant species have evolved sophisticated immune systems to defend against infectious agents. Still, infections are debilitating and sometimes fatal [2]. One way to stay healthy is to avoid contact with pathogens, and many species engage in a wide range of prophylactic behaviors against infections [3]. This behavioral defense against infection includes perceptual mechanisms for detecting cues of infectious pathogens which, in turn, trigger disgust (a basic emotion which facilitates avoidance of infectious disease and toxins), and behavioral avoidance [4]. Still, our current understanding of the human behavioral repertoire that reduces the risk of infection remains limited. Notably, humankind is an inherently social species and avoidance of infected individuals may be costly as it deprives individuals from valuable social interactions such as mating opportunities, social learning, childcare, or foraging. Hence, pathogen avoidance behaviors should only be adopted when their expected benefits outweigh their expected costs. However, the characteristics that influence how an individual respond to pathogens remains to be determined.

Many social species, including mandrills, mice, and lobsters, identify sick conspecifics via olfactory and visual cues and strategically avoid physical contact with them [3,5–7]. Humans are likewise able to identify sick individuals based on visual cues of sickness, such as vomiting, coughing, sneezing or rashes, and following exposure to these cues, individuals likely

feel disgust and engage in behavioural avoidance (e.g., increase their interpersonal distance to the sick conspecific) [8,9]. However, these aforementioned sickness cues, are typically expressed when people have been contagious for a while and at the later stages of infection. The ability to detect and avoid sick individuals in early phases of an infection would be even more efficient in reducing the likelihood of contamination.

A growing body of research suggests that humans can detect cues characterizing the early phases of sickness. For instance, a series of studies that experimentally-induce sickness via injection with an endotoxin (lipopolysaccharide, LPS) showed that sickness can be identified above chance already a couple of hours after injection based on faces, body odors and biological motion [10–14]. Skin coloration is likely a reliable cue of early sickness as naïve observers perceive sick volunteers as paler in their skin and lip color compared to the same volunteers when healthy [10]. In fact, this change in skin color is objectively measurable as early as one hour after injection with endotoxin [15]. Further, sick volunteers exhibit more hanging eyelids and drooping corners of the mouth, and are rated as less attractive, less healthy, and expressive of more negative emotions compared to healthy faces [10,12,16]. Finally, there is evidence that sick volunteers are less liked, suggesting that cues of sickness may trigger adaptive social responses, such as avoidance [12,13].

Despite the evidence that sickness cues can be perceived by observers, and that such perception leads to attitudes that may precede disease-avoiding behaviors, little is known about inter-individual variation in the detection of these cues, or how this variation influences the propensity for behavioral avoidance. There is evidence that some personality traits are related to individual reactions to disease-related stimuli. For example, individuals with higher perceived vulnerability to disease seem to be more alert towards conspecifics who display a sickness cue, or

towards objects that could potentially transmit an infection, and upon detection, they may exhibit stronger avoidance behaviors [4,17]. Moreover, individuals suffering from health anxiety, a persistent fear of being or falling ill, perceive others as less healthy, and also rate the risk of contagion as greater compared to individuals with low levels of health anxiety [18]. In parallel, individuals that are easily disgusted tend to exhibit more behavioral avoidance of sickness connoting stimuli [19]. To our knowledge, a few studies have examined the influence of such personality traits on individual reactions to early cues of sickness, none of which found any associations [13,20]. However, these studies used relatively small sample sizes (N=77 and 44, respectively), preventing any firm conclusions.

Sickness detection and behavioral avoidance may differ between men and women. While parental investment is a crucial resource for human children, paternal investment is facultative and shows larger inter-individual variation compared to maternal investment [21,22]. For example, evidence from (mostly) natural fertility populations suggests that while the presence of a mother is pivotal for child survival, fathers have little effect on child survival [22]. Hence, women may be more likely to avoid sick conspecifics than men, because they have a more central role in protecting themselves and offspring from disease [9,23]. Moreover, as women are more susceptible to sexually transmitted infections, heightened perception and response to sickness cues may be advantageous, in reducing their infection risk [23,24]. As such, women may display a greater accuracy – and a lower bias – in detecting disease-relevant cues, or, alternatively, an increased likelihood in falsely identifying non-infectious individuals as infectious (i.e. a higher bias). In line with this, women report higher disgust sensitivity when exposed to conspicuous disease-related stimuli [9,25,26] and exhibit higher scores on the Disgust Scale, a scale sampling seven domains of disgust elicitors [27,28], as compared to men. However, whether women are

more accurate (i.e., less biased) or less accurate (i.e., more biased) in detecting early and more subtle cues of sickness remains unknown.

In the present study, we tested whether humans can discriminate between “healthy” and “sick” faces, and whether the perception of cues of sickness is associated with sex, self-reported health, and disease-related personality traits, by using signal detection methodology that enables the separation of detection accuracy from the behavioral tendency (bias) to categorize stimuli as disease cues [29]. We used a unique set of facial photographs from an LPS injection study. The set consists of pairs of photographs for a number of individuals: one taken in a healthy control condition, and the other in an experimental sickness condition. In a large online study, we presented these facial photographs to participants (N=683) sequentially and asked them to identify each face as either sick or healthy. They also completed several disease-related questionnaires. Based on previous studies [10,12,13], we predicted that humans would be better than chance at discriminating between healthy and sick individuals based solely on early facial cues of sickness. We also predicted that perception of disease cues would differ between men and women, with women being more biased – i.e., more likely to overestimate the presence of sickness cues in a face. In addition, we explored if women were worse or better at differentiating between sick and healthy faces compared to men. Finally, we predicted that individuals who are more sensitive to disgust, vulnerable to disease, concerned about their health, and self-report a poorer health, would also be more likely to overestimate the presence of facial sickness cues.

## 2. Methods

The design and analysis plan were preregistered on OSF prior to data collection (<https://doi.org/10.17605/OSF.IO/P85EZ>).

### 2.1. Facial stimuli

#### 2.1.1 Collection

The facial stimuli (photographs) were collected in a previous study investigating sickness detection and behavioral and physiological changes during acute systemic inflammation. A more detailed description of the experimental model and stimuli acquisition is available in [10,30]. The initial study, as well as other studies that used the stimuli, were approved by the regional ethical review board in Stockholm, Sweden (Dnr 2014/1946-31/1). In short, twenty-two healthy volunteers (mean age 23 years; 9 women) participated in a within-subjects, double-blind and placebo-controlled study using the experimental endotoxemia model. These participants, hereafter called donors, were randomly assigned to either first receive a lipopolysaccharide (LPS) injection (Escherichia coli endotoxin, Lot HOK354, CAT number 1235503, United States Pharmacopeia, Rockville, MD, USA) at 2.0 ng/kg body weight or a placebo injection (0.9% NaCl), and the reverse treatment approximately one month later. In both conditions (LPS and Placebo), pro-inflammatory cytokines, subjective sickness ratings, and tympanic temperature were measured to confirm that a significant inflammatory response was caused by the LPS administration (for more details see [30]). Facial photographs of the donors were taken two hours after injection, which corresponds, in the LPS condition, to the approximate peak of the pro-inflammatory cytokine levels [10,30]. The facial photographs were obtained using a Nikon D90 (Nikon Corp., Tokyo, Japan), with a focal length of 50 mm (a 50 mm lens was used), aperture speed at 1/125, and 200 ISO and the



same lighting conditions. All donors wore identical white t-shirts, were asked to sit comfortably, look straight into the camera, and express a neutral facial expression, and were restricted from wearing any make-up.

### 2.1.2 Selection and preparation of facial stimuli

To prevent recognition of donors, and circulation of the donors' faces on the internet, we created realistic virtual composite faces based on the donors facial photographs that kept sickness cues intact [31]. All composite faces were made with *Webmorph*, a web-based version of *Psychomorph* and this involved three steps [32]. First, same-sex donors were randomly assigned to pairs. Second, the two paired faces were combined within a given condition (LPS or Placebo) resulting in composites presenting a 50:50 combination of shape, color, and texture information from both original faces. Third, each composite was cropped to include the face only and the background was changed to black. This procedure resulted in a new dataset of 20 facial composites (10 healthy and 10 sick) from 6 virtual men and 4 virtual women (for an illustration of our facial stimuli, please refer to Figure 1, which depicts a composite of the faces of the 20 donors in each condition.) Out of the 22 donors, one man and one woman were excluded because their faces were too obscured by hair, making the composites appear unrealistic.

## 2.2 Main experiment

### 2.2.1 Participants

We collected full data sets from 343 women (Mean Age = 34.30 years, SD = 12.18, range = 18 to 72 years) and 340 men (Mean Age = 35.29, SD = 12.95, range = 18 to 75 years). Age distribution of the participants is presented in Figure S1 of the supplementary material (SM). Another 32 participants started the study but were excluded because they failed the quality check questions

(see below). Our sample exceeded our initial aim (N=680), which was based on a power analysis allowing for detection of a weak association between personality traits and perception of facial cues of sickness ( $r = 0.15$ ) with a power of 80% and an alpha level = 0.001 for a planned directional hypothesis, or 80% power to detect a sex difference as small as  $d = 0.22$  with an alpha level of 0.05 (two-directional hypothesis). All participants were recruited using the online recruitment system *Prolific* (<https://www.prolific.co>). Participants whose first language was English (the language used during the experimental procedure), and who identified as either male or female (both biological sex and gender) according to their prolific profile were invited to the study. All participants received a flat fee of £2.50 (currency used by the *Prolific* platform) as compensation for their participation. They completed the study in 11 (SD = 6) minutes on average.

### 2.2.2 Procedure

After consenting to participate in the study, participants performed a discrimination task and then completed several questionnaires on disgust sensitivity to body odors, health anxiety, and perceived vulnerability to disease. Participants could withdraw their participation at any time in which case their data were not saved. The online procedure was programmed with *Psychopy 3.0* [33]. The discrimination task and questionnaires are described below.

### 2.2.3 Discrimination between sick and healthy individuals

The task was to discriminate between sick and healthy facial composites. Participants were not informed that the facial photographs were composites rather than individual faces. They were instructed to look at a series of 20 composite faces presented one at a time, and to push one of two buttons to indicate whether each face depicted a sick or healthy individual. Within each sex, half of the participants were instructed to push the left-arrow button for healthy and right-arrow for

sick, while the other half were assigned the opposite pairing. The facial composites were displayed for 5 seconds, each in a random order. Participants could indicate their choice at any time (no time-out), including after the face had disappeared. Before the task, participants completed two practice trials with a male and a female healthy facial composite (order randomized) created from faces that were not used in the experiment.

#### 2.2.4 Disease-related questionnaires

After the identification task, participants completed several short questionnaires. The first questionnaire, the Body Odor Disgust Scale (BODS), is a 12-item self-report measure that assesses feelings of disgust towards body-related odors such as upper body sweat and urine, with two subscales: disgust for one's own body odors (internal sources) and those of other people (external sources) [34]. Participants rated the extent to which each scenario elicited disgust on a Likert type of scale ranging from 1 (not disgusting at all) to 5 (extremely disgusting). The second questionnaire, Perceived Vulnerability to Disease (PVD), is a 15-item self-report tool assessing individuals' concerns of disease transmission with two subscales: the germ subscale, which targets discomfort in situations with potential transmission of pathogens, and the infectability subscale, which measures perceived susceptibility to infectious diseases [35]. Because the last item of the questionnaire is outdated ("I avoid using public telephones because of the risk that I may catch something from the previous user"), it was replaced with "I avoid using public telephones or grabbing handrails in public transport because of the risk that I may catch something from the previous user". Participants responded to each item on a scale ranging from 1 (strongly disagree) to 7 (strongly agree). The third questionnaire used was the 14-item Short Health Anxiety Inventory (SHAI), which measures clinical health anxiety [36]. Each of the 14-items consists of a group of four statements that are scored from 0 to 3. Participants had to select the statement that described

their feelings the best. For each questionnaire, an individual score was calculated as the mean across items. Finally, they were asked to rate their health on a scale ranging from 1 (very bad) to 5 (very good). Covid-related questions were also asked, but these questions were not analyzed in the current study.

### 2.2.5 Exclusion criteria

We included a control question in both the BODS and PVD questionnaires to confirm that the participants were following the instructions and not giving random responses. For the control question in the BODS questionnaire, participants were instructed to not click on the scale but rather on a small cross displayed at the bottom of the screen. For the control question in the PVD questionnaire, participants were instructed to select the answer “strongly agree”. Participants who responded incorrectly on any of the control questions were excluded (n=32).

## 2.3. Statistical analyses

All analyses follow the preregistered analysis plan unless otherwise specified. Descriptive statistics including means, SD, and t-test results comparing men's and women's reaction time, self-reported health, BODS, PVD, and SHAI are displayed in Table S1 in SM.

### 2.3.1 Discrimination between sick and healthy individuals

To determine whether human observers could discriminate between sick and healthy individuals based on early facial cues of sickness, we first assessed discrimination accuracy, using the method of signal detection. We calculated both the discriminability ( $d'$ ), from the probability of true sickness detections and false alarm, and the decision criterion ( $c$ ), for each participant (package *psycho* in R). While the discriminability  $d'$  indicates how well sick and healthy faces can be

discriminated (with  $d' = 0$  indicating no discrimination at all [random choice]) and higher values indicating an increased ability), the decision criterion reflects how biased the participant is from a liberal ( $c < 0$ , most of the faces are considered sick) to a conservative bias ( $c > 0$ , few faces are considered sick). We used a one-sided one-sample Wilcoxon signed rank test (instead of the preregistered one-sided one-sample Student t-test to account for the non-gaussian distribution of the data) to examine whether the mean of the discriminability  $d'$  was significantly above 0.

We then examined whether sickness condition improved participants' ability to discriminate between sick and healthy facial composites by using a Generalized Linear Mixed Model (GLMM) with a binomial error structure (*glmer* function of the *lme4* R package). Our dependent variable was the decision made by a participant (0 when they answered that the face was healthy, 1 when answering sick) for each of the 20 facial composites. Our explanatory variable, sickness condition, represented the two categories of the composite (healthy vs. sick). We included a random intercept for each participant's and facial composite's ID, and random slopes for condition by participant and composite.

### 2.3.2 Sex difference and the perception of sickness

To determine whether women overestimate the presence of sickness cues (i.e., classified more faces as sick) compared to men, we used a one-sided Wilcoxon rank sum test with continuity correction (instead of the preregistered one-sided Student t-test to account for the non-gaussian distribution of the data) testing whether women's decision criterion  $c$  was significantly lower than men's decision criterion.

We also performed exploratory analyses examining potential sex differences in the accuracy of discrimination between sick and healthy facial composites. We first used two-sided Wilcoxon rank sum tests with continuity correction comparing the percentage of correct identifications and the discriminability ( $d'$ ) between men and women.

Given that sex differences related to age, reaction time, general health, or disease-related personality traits, might potentially account for the higher sickness discrimination rate in women, we controlled for the robustness of the results by conducting multiple linear regression analyses. In these analyses, we used  $c$ ,  $d'$ , and the percentage of correct identifications as dependent variables, while including sex as an exploratory variable and age, reaction time, self-rated health, PVD, BODS, and SHAI as confounding variables. Because these variables may be intercorrelated, we checked for multicollinearity among the confounding variables by calculating the variance inflation factor (VIF). However, we found no evidence of multicollinearity, as all variables showed a low VIF value for all models ( $VIF < 1.38$ ).

Finally, we used a hierarchical drift diffusion modeling (HDDM) approach. One of the advantages of this approach is that it allows for the combination of both reaction times and accuracy into one parameter – drift rate ( $v$ ). The drift rate reflects how fast a person can accumulate information that leads to a correct decision. Higher  $v$  values reflect better performance in a shorter time. We built a model using sickness condition (healthy vs. sick facial composites), participants' sex and the interaction between sex and condition as exploratory variables. Using the *hddm* Python package [37], we estimated a model in which the drift rate was specified to vary for men and women, separately for the sick and healthy facial composite conditions. We report the estimated drift rate  $v$  along with its corresponding 95% credibility intervals. We used a Markov Chain Monte Carlo (MCMC) sampling method to draw 10000 posterior samples with a burn-in of 2000 samples, and

evaluated the model fit by visually inspecting the chain traceplots and autocorrelation plot for each model parameter (i.e., threshold/decision boundary  $a$ , the non-decision time constant  $t$ , and  $v$ ). Additionally, we reran the same model five times with chains of 2000 samples (500 burn-in) to obtain and compare the Gelman-Rubin statistic, which allowed us to judge chain stability [38].

### 2.3.3 Disease-related traits and perception of sickness cues

To investigate whether participants with poorer health, who are more sensitive to disgust, vulnerable to disease, and/or concerned about their health, overestimate the presence of sickness cues (i.e., have a more liberal bias), we used a linear regression examining whether a participant's decision criterion ( $c$ ) was negatively associated with self-reported health and the score for each of the three questionnaires (BODS, PVD, SHAI) and their subscales. Because women self-reported higher scores than men on all the questionnaires [34,35], we controlled for the robustness of the results by re-running the same models with the participant's sex as confounding variable.

Statistical analyses were performed with R, version 4.1.3 [39]. For the linear regressions and GLMMs, the statistical significance of each variable was tested with likelihood ratio tests comparing the full model to those without the term of interest.

## 3. RESULTS

### 3.1 Discrimination between sick and healthy individuals

In line with our prediction, human observers could discriminate between sick and healthy facial composites; on average, participants correctly identified  $60.3 \pm 9.9\%$  of the faces (chance level 50%, Figure 2A), and the discriminability  $d'$  was significantly above zero ( $d' = 0.56 \pm 0.53$ ; one-sided Wilcoxon signed rank test with continuity correction:  $V = 160662$ ,  $p < .0001$ , effect size  $r =$

0.76, Figure 2B). Similarly, the GLMM indicated that condition significantly influenced participants' decisions ( $\beta = 1.22$ ,  $SE = 0.17$ ,  $X^2 = 50.28$ ,  $df = 1$ ,  $p < .0001$ ).

### 3.2 Sex difference

Contrary to our hypothesis, women did not report the presence of sick faces (decision criterion  $c$ , Mean  $\pm$  SD =  $0.42 \pm 0.50$ ) significantly more than men ( $0.40 \pm 0.49$ ; one-sided Wilcoxon rank sum test with continuity correction:  $W = 60538$ ,  $p = .81$  effect size  $r = 0.03$ ). However, our exploratory analyses indicate that women were better than men at differentiating between sick and healthy faces, as indicated by both percentage of correct identifications (Women =  $61 \pm 1\%$ , Men =  $59 \pm 1\%$ , Figure 2A) and discriminability  $d'$  (Women =  $0.62 \pm 0.5$ , Men =  $0.50 \pm 0.52$ , Figure 2B) being significantly higher for women (correct identifications: two-sided Wilcoxon rank sum test with continuity correction:  $W = 64878$ ,  $p = .01$ , effect size  $r = 0.10$ ; Discriminability  $d'$ :  $W = 65517$ ,  $p = .005$ ,  $r = 0.11$ ). These findings remained robust even when controlling for confounding variables such as age, BODS, PVD, HAI, self-reported health, and reaction time, with participants' sex significantly influencing percentage of correct identifications ( $\beta = -0.02$ ,  $SE = 0.01$ ,  $F = 6.87$ ,  $df = 1$ ,  $p = .009$ , Table S2 in SM) and discriminability  $d'$  ( $\beta = -0.11$ ,  $SE = 0.04$ ,  $F = 7.58$ ,  $df = 1$ ,  $p = .006$ , Table S2) but not decision criterion  $c$  ( $\beta = -0.01$ ,  $SE = 0.04$ ,  $F = 0.04$ ,  $df = 1$ ,  $p = .84$ , Table S2).

In addition, the drift-diffusion analysis demonstrates that women were not only more accurate but are also faster than men at differentiating between sick and healthy faces (Figure 3). Indeed, women exhibited higher drift rates, namely accumulated more evidence of



sickness/healthiness per time unit, than men for both sick (women:  $\nu = -0.1$  [-0.17, -0.04] vs. men:  $\nu = -0.14$  [-0.2, -0.08]) and healthy faces (women:  $\nu = 0.76$  [0.7, 0.83] vs. men:  $\nu = 0.64$  [0.58, 0.7]), with the probability of women having a higher drift rate than men being 0.78 for sick and 0.99 for healthy faces. Chains had symmetrical traceplots and low autocorrelation, and the R-hat Gelman-Rubin statistic was  $1 \pm 0.02$  (Mean  $\pm$  SD) indicating that the model converged well.

### 3.3 Influence of disease-related traits

Next, we tested whether participants self-reported health, disgust sensitivity towards body odors, vulnerability to disease, and health anxiety were negatively associated with the decision criterion, a lower decision criterion reflecting a more liberal bias (i.e., most of the faces are considered sick). In line with our prediction, we found that individuals who are more easily disgusted by body odors, as measured by the Body Odor Disgust Scale (BODS), are more sensitive to potential sickness cues (liberal bias), i.e., classify more faces as sick ( $\beta = -0.10$ , SE = 0.03,  $F = 10.12$ ,  $df = 1$ ,  $p < .002$ , Figure 4A). This relationship seems mainly driven by disgust feelings towards *others'* body odors (i.e., external body odor source subscale,  $\beta = -0.11$ , SE = 0.03,  $F = 15.54$ ,  $df = 1$ ,  $p < .0001$ , Figure 4B) compared to their own body odors (i.e., internal body odor source subscale,  $\beta = -0.05$ , SE = 0.03,  $F = 3.16$ ,  $df = 1$ ,  $p = .08$ , Figure 4C). These results were robust to the inclusion of participant sex as a confounding variable - both disgust sensitivity towards body odors (BODS total score,  $\beta = -0.10$ , SE = 0.03,  $F = 10.24$ ,  $df = 1$ ,  $p < .002$ ) and towards *others'* body odors (external body odor source sub-scale,  $\beta = -0.11$ , SE = 0.03,  $F = 15.88$ ,  $df = 1$ ,  $p < .0001$ ) remained significantly and negatively related to the decision criterion, and the

relationship between internal body odor source score and the decision criterion remained non-significant ( $\beta = -0.05$ ,  $SE = 0.03$ ,  $F = 3.14$ ,  $df = 1$ ,  $p = .08$ ). Contrary to our prediction, however, a participant's decision criterion was not significantly associated with self-reported health ( $\beta = 0.01$ ,  $SE = 0.02$ ,  $F = 0.42$ ,  $df = 1$ ,  $p = .52$ ), perceived vulnerability to disease (PVD; total score:  $\beta = -0.02$ ,  $SE = 0.02$ ,  $F = 0.96$ ,  $df = 1$ ,  $p = .33$ ; germ subscale:  $\beta = -0.003$ ,  $SE = 0.02$ ,  $F = 0.03$ ,  $df = 1$ ,  $p = .86$ ; infectability subscale:  $\beta = -0.02$ ,  $SE = 0.02$ ,  $F = 1.57$ ,  $df = 1$ ,  $p = .21$ ), nor with the health anxiety (SHAI;  $\beta = -0.05$ ,  $SE = 0.04$ ,  $F = 1.42$ ,  $df = 1$ ,  $p = .23$ ).

#### 4 – DISCUSSION

Because of the cost of sociality in term of transmission of infectious agents, it has been hypothesized that social animals evolved a complementary behavioral defense strategy [3,4]. This prophylactic defense strategy includes the ability to accurately identify sickness cues, and a motivational response to avoid interactions with potentially infected individuals [4]. In fact, there is evidence that social species, including humans, identify sick conspecifics based on cues from different sensory modalities such as faces, body odors or biological motion [5–7,10–12,14]. Interestingly, humans do not only detect conspicuous cues of sickness (e.g., coughs, rashes, or sneezes) but also cues expressed only 2 hours after the onset of an induced systemic inflammatory response [11–13]. Here, we replicate these findings by showing that our participants could discriminate between sick and healthy individuals based on facial composites, again supporting the view that humans are able to detect sickness cues only a couple of hours after exposure to an endotoxin stimulus.

The large sample of participants also provided enough power to investigate whether sensitivity to early facial sickness cues was associated with sex and disease-related personality traits. While we could not support the hypothesis that women have a more liberal decision criterion compared to men, exploratory analyses showed that women were faster and slightly better at discriminating between sick and healthy individuals. We hypothesized that the perception of sickness cues would differ between men and women, due to the evolutionary pressure on women to protect both themselves and their offspring from infections disease [9] and because women exhibit substantially higher levels of disgust than men [23,27,40], including disgust towards pathogen connoting stimuli [9,25,26,28,34]. Better sickness detection and greater disgust sensitivity could enable women to reduce contaminations by triggering behavioral avoidance towards sick individuals. While the difference in discrimination accuracy between men and women is rather small in terms of effect size ( $r = .10$ ), a hierarchical drift diffusion modeling approach showed that by combining both reaction times and accuracy, the likelihood of women performing better than men in a shorter period of time is actually over 78%. Besides, in our identification task, participants could base their decision through only one sensory cue. Hence, it still remains plausible that the sex difference may change – i.e., be greater or smaller – when individuals have the possibility of integrating sickness cues from different sensory modalities, such as body odors [11,13,20], voice [41], or gait [14]. In line with this, identification of sickness has been more accurate when visual and olfactory cues are presented simultaneously [12,13]. Finally, a small sex difference in sickness detection could possibly be related to a greater sex difference in behavioral avoidance; thus, further investigations should explore whether sickness cues trigger behavioral avoidance and whether behavioral avoidance differs between men and women.

There are several alternative explanations for the demonstrated sex difference in ability to discriminate between healthy and sick faces. For example, women might have been more motivated than men to carry out the task to their best ability. However, such an effect would typically be accompanied by longer inspection time and thus manifest as a speed-accuracy trade-off. Instead, women were not only more accurate but also faster at responding. Differential exposure to sickness due to women's higher parental care investment as compared to men [21,22] could also lead to a better ability to discriminate between healthy and sick individuals through a learning process. In addition, many studies indicate that women outperform men in tasks related to face recognition and processing of facial emotional expressions [42,43]. This higher ability to read facial expressions could be particularly important in identifying sick individuals as sick faces display more negative emotions such as sadness and disgust, and less happiness and surprise [16]. Therefore, it is possible that the higher accuracy of women in discriminating healthy from sick faces observed in our study stems from their better ability to identify facial expressions. Finally, another explanation could be the higher prevalence of color blindness in men (around 8% in Caucasian Europeans) as compared to women (0.4%) [44]. Previous studies found that skin coloration is an important facial cue of sickness: after endotoxin injection, facial skin become lighter and less red [15] and paleness of facial skin and lips were found to be associated with apparent sickness [10]. Although other cues, unrelated to skin color, were also associated with apparent sickness (such as having a more swollen face, droopier corners of the mouth, and more hanging eyelids) [10,16], color blindness could impede the ability to discriminate sick from healthy faces. We estimated the potential influence of color blindness in our data set in an additional analysis where we simulated an inability to detect sickness due to color blindness in 8% of the females due to color blindness, matching the male prevalence. With this background, we randomly

selected 7.6% (n=26) of the females and replaced their decisions by random responses (50% sick and 50% healthy responses). In this conservative analysis, women's  $d'$  was still higher ( $d' = 0.57$ ) than men's ( $d' = 0.50$ ; average of 1000 iterations), suggesting that color blindness is not likely to be the only factor explaining women's higher accuracy. This exploratory result should nonetheless be replicated in a study controlling for color blindness.

In line with our prediction, we found that individuals who are more disgust sensitive to body odors were more prone to overestimate the presence of sickness cues and that this effect was specifically driven by disgust toward others' body odors. However, no associations were found with perceived vulnerability to disease or health anxiety. It has been argued that disgust is an evolved motivational system for disease-avoidance [40]. Indeed, pathogen-connoting stimuli trigger high levels of disgust [9,25,26,28,34]. The perception of others' body odors indicates an interpersonal proximity in which pathogens have an increased chance of being transmitted, and diseases are known to affect the body odors of sick individuals including the smell of their sweat, urine, and breath [45]. High BODS scores are positively associated with disgust ratings of sweat odors suggesting that BODS is a valid marker of olfactory disgust [46]. Our finding that body odor disgust predicts the tendency to classify individuals as sick, but not discriminability, may be linked with previous research showing that individuals that are highly disgusted by body odors have more negative views of refugees [47,48] and sympathize more with right-wing authoritarianism [49], two ideological dispositions that may restrict inter-group contact and social mobility. Hence, it could be speculated that high body odor disgust sensitivity arises from a dispositional bias that leans toward classifying other individuals as pathogen threats.

## Study limitations

We used a database of early cues of sickness stimuli, collected during a study that used a within-subjects placebo-controlled experimental endotoxemia model, in the largest study on sickness detection performed thus far with almost 700 participants. However, our study presents a few limitations. First, our results are limited to visual detection and sickness stimuli that reflect a systemic inflammation (i.e., a non-specific immune reaction to endotoxins). Akin to body odors for which specific smells may be associated with specific diseases [45], some diseases may be detected from different visual cues, a trivial example being chicken pox. Further studies are thus necessary to generalize our findings to different naturally occurring infectious diseases in a context more ecologically valid. Second, because of the COVID-19 pandemic, we conducted this experiment online rather than in a controlled laboratory environment. Although we used two quality check questions, it remained difficult to assess participants' motivation to succeed at the task, as well as their honesty when completing the questionnaires. However, a recent published study comparing several recruitment platforms for online behavioral research found that *Prolific* provides high-quality data regarding participants' attention, comprehension, honesty, and reliability [50]. Third, conducting the study online forced us, due to ethical considerations, to use facial composites rather than individual faces. Facial composites were created by averaging real facial photographs; healthy composites were obtained by averaging individuals in a healthy state, and sick composites were obtained by averaging individuals who underwent an induced inflammatory sickness response. Although this method allows for creating realistic faces, we do not know if the participants realized that the faces were morphs (no participant commented on this aspect) and how this method might have affected the results. The discriminability index ( $d' = 0.56$ ) was, nevertheless, almost identical ( $d' = .57$  calculated based on their reported sensitivity [52]) and

specificity [.30]) to that found in a previous study using the original photos [10], indicating that the relevant sickness cues were preserved. Furthermore, it is still unclear whether the ability to distinguish between sick and healthy individuals relies on recognizing specific facial cues of sickness, noticing the absence of certain facial cues of healthiness, or identifying general deviations in facial traits, such as emotional expressions or signs of fatigue, that are associated with sickness in the particular task used in the study. This aspect was not explored here, and it is still to be determined. Finally, it is important to note that our facial database of sickness cues is limited to Swedish participants, and the raters we recruited were only native English speakers, which limits the generalization of our findings to other cultures.

In conclusion, our study provides support that humans can differentiate between sick and healthy individuals based solely on early facial cues of sickness and suggests that, while women did not overestimate the presence of sickness cues, women were significantly, albeit only slightly, better at discriminating between sick and healthy individuals than men. In addition to replicating this exploratory finding, future research should investigate whether women's putative superior performance is related to increased avoidance of sick individuals and whether our findings generalize to naturally occurring contagious diseases as well as to other populations.

## ACKNOWLEDGEMENTS

We express our gratitude to Megan N. Williams and Supreet Saluja for their English copy-editing assistance, as well as to Julie Lasselin for her contribution to the data collection of the facial stimuli.

## AUTHOR CONTRIBUTIONS

MO and AT developed the study concept and designed the methodology. AT created the virtual stimuli, the computer-based experiment, recruited the participants, and collected data. ET, MZ, and AT conducted formal analysis. AT drafted the manuscript. ET, MO, MZ, JO, JA, and ML provided critical revisions. All authors approved the final version of the manuscript for submission.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available on OSF at <https://doi.org/10.17605/OSF.IO/P85EZ>.

## FUNDING

The behavioral experiment in the present paper was supported by Swedish Research Council Grants 2016-02742 (MO), 2020-02567 (MO), and 2021-03184 (AT) and Swedish Foundation for Humanities and Social Sciences Grant P12-1017 (MO).

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.



## REFERENCES

1. Wu X, Lu Y, Zhou S *et al.* Impact of climate change on human infectious diseases: Empirical evidence and human adaptation. *Environ Int* 2016;**86**:14–23.
2. Shakhar K, Shakhar G. Why do we feel sick when infected—can altruism play a role? *PLoS Biol* 2015;**13**:e1002276.
3. Hart BL. Behavioural defences in animals against pathogens and parasites: parallels with the pillars of medicine in humans. *Philos Trans R Soc B Biol Sci* 2011;**366**:3406–17.
4. Schaller M, Park JH. The behavioral immune system (and why it matters). *Curr Dir Psychol Sci* 2011;**20**:99–103.
5. Arakawa H, Cruz S, Deak T. From models to mechanisms: Odorant communication as a key determinant of social behavior in rodents during illness-associated states. *Pioneer Res Affect Neurosci Celebr Work Dr Jaak Panksepp* 2011;**35**:1916–28.
6. Behringer DC, Butler MJ. Disease avoidance influences shelter use and predation in Caribbean spiny lobster. *Behav Ecol Sociobiol* 2010;**64**:747–55.
7. Poirotte C, Massol F, Herbert A *et al.* Mandrills use olfaction to socially avoid parasitized conspecifics. *Sci Adv* 2017;**3**:e1601721.
8. Curtis V. Infection-avoidance behaviour in humans and other animals. *Trends Immunol* 2014;**35**:457–64.
9. Curtis V, Aunger R, Rabie T. Evidence that disgust evolved to protect from risk of disease. *Proc R Soc Lond B Biol Sci* 2004;**271**:S131–3.
10. Axelsson J, Sundelin T, Olsson MJ *et al.* Identification of acutely sick people and facial cues of sickness. *Proc R Soc B Biol Sci* 2018;**285**:20172430.
11. Olsson MJ, Lundström JN, Kimball BA *et al.* The scent of disease: human body odor contains an early chemosensory cue of sickness. *Psychol Sci* 2014;**25**:817–23.
12. Regenbogen C, Axelsson J, Lasselin J *et al.* Behavioral and neural correlates to multisensory detection of sick humans. *Proc Natl Acad Sci* 2017;**114**:6400–5.
13. Sarolidou G, Axelsson J, Lasselin J *et al.* People expressing olfactory and visual cues of disease are less liked. *Philos Trans R Soc B Biol Sci* 2020:20190272.
14. Sundelin T, Karshikoff B, Axelsson E *et al.* Sick man walking: Perception of health status from body motion. *Brain Behav Immun* 2015;**48**:53–6.
15. Henderson AJ, Lasselin J, Lekander M *et al.* Skin colour changes during experimentally-induced sickness. *Brain Behav Immun* 2017;**60**:312–8.
16. Sarolidou G, Axelsson J, Sundelin T *et al.* Emotional expressions of the sick face. *Brain Behav Immun* 2019;**80**:286–91.
17. Ackerman JM, Becker DV, Mortensen CR *et al.* A pox on the mind: Disjunction of attention and memory in the processing of physical disfigurement. *J Exp Soc Psychol* 2009;**45**:478–85.

18. Hedman E, Lekander M, Karshikoff B *et al.* Health anxiety in a disease-avoidance framework: Investigation of anxiety, disgust and disease perception in response to sickness cues. *J Abnorm Psychol* 2016;**125**:868.
19. Fan Q, Olatunji BO. Individual differences in disgust sensitivity and health-related avoidance: Examination of specific associations. *Personal Individ Differ* 2013;**55**:454–8.
20. Sarolidou G, Tognetti A, Lasselin J *et al.* Olfactory Communication of Sickness Cues in Respiratory Infection. *Front Psychol* 2020;**11**:1004.
21. Geary DC. Evolution and proximate expression of human paternal investment. *Psychol Bull* 2000;**126**:55–77.
22. Sear R, Mace R. Who keeps children alive? A review of the effects of kin on child survival. *Evol Hum Behav* 2008;**29**:1–18.
23. Al-Shawaf L, Lewis DMG, Buss DM. Sex Differences in Disgust: Why Are Women More Easily Disgusted Than Men? *Emot Rev* 2018;**10**:149–60.
24. Tybur JM, Lieberman D, Griskevicius V. Microbes, mating, and morality: individual differences in three functional domains of disgust. *J Pers Soc Psychol* 2009;**97**:103.
25. Prokop P, Usak M, Fančovičová J. Health and the avoidance of macroparasites: a preliminary cross-cultural study. *J Ethol* 2010;**28**:345–51.
26. Al-Shawaf L, Lewis DM. Exposed intestines and contaminated cooks: Sex, stress, & satiation predict disgust sensitivity. *Personal Individ Differ* 2013;**54**:698–702.
27. Haidt J, McCauley C, Rozin P. Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personal Individ Differ* 1994;**16**:701–13.
28. Olatunji BO, Williams NL, Tolin DF *et al.* The Disgust Scale: item analysis, factor structure, and suggestions for refinement. *Psychol Assess* 2007;**19**:281.
29. Macmillan NA, Creelman CD. *Detection Theory: A User's Guide*. Psychology press, 2004.
30. Lasselin J, Treadway MT, Lacourt TE *et al.* Lipopolysaccharide alters motivated behavior in a monetary reward task: a randomized trial. *Neuropsychopharmacology* 2017;**42**:801–10.
31. Bovet J, Tognetti A, Pollet TV. Methodological issues when using face prototypes: A case study on the Faceaurus dataset. *Evol Hum Sci* 2022;**4**:e48.
32. Tiddeman B, Burt M, Perrett D. Prototyping and transforming facial textures for perception research. *IEEE Comput Graph Appl* 2001;**21**:42–50.
33. Peirce J, Gray JR, Simpson S *et al.* PsychoPy2: Experiments in behavior made easy. *Behav Res Methods* 2019;**51**:195–203.
34. Liuzza MT, Lindholm T, Hawley C *et al.* The Body Odor Disgust Scale (BODS): Development and Validation of a Novel Olfactory Disgust Assessment. *Chem Senses* 2016;**42**:499–508.
35. Duncan LA, Schaller M, Park JH. Perceived vulnerability to disease: Development and validation of a 15-item self-report instrument. *Personal Individ Differ* 2009;**47**:541–6.

36. Salkovskis PM, Rimes KA, Warwick HMC. The Health Anxiety Inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. *Psychol Med* 2002;**32**:843.
37. Wiecki T, Sofer I, Frank M. HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Front Neuroinformatics* 2013;**7**:14.
38. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci* 1992:457–72.
39. R Core Team. R: a language and environment for statistical computing. 2022.
40. Oaten M, Stevenson RJ, Case TI. Disgust as a disease-avoidance mechanism. *Psychol Bull* 2009;**135**:303.
41. Milling M, Pokorny FB, Bartl-Pokorny KD *et al*. Is Speech the New Blood? Recent Progress in AI-Based Disease Detection From Audio in a Nutshell. *Front Digit Health* 2022;**4**.
42. McClure EB. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychol Bull* 2000;**126**:424.
43. Olderbak S, Wilhelm O, Hildebrandt A *et al*. Sex differences in facial emotion perception ability across the lifespan. *Cogn Emot* 2019;**33**:579–88.
44. Birch J. Worldwide prevalence of red-green color deficiency. *JOSA A* 2012;**29**:313–20.
45. Shirasu M, Touhara K. The scent of disease: volatile organic compounds of the human body related to disease and disorder. *J Biochem (Tokyo)* 2011;**150**:257–66.
46. Liuzza MT, Olofsson JK, Sabiniewicz A *et al*. Body Odor Trait Disgust Sensitivity Predicts Perception of Sweat Biosamples. *Chem Senses* 2017;**42**:479–85.
47. Zakrzewska M, Olofsson JK, Lindholm T *et al*. Body odor disgust sensitivity is associated with prejudice towards a fictive group of immigrants. *Physiol Behav* 2019;**201**:221–7.
48. Zakrzewska MZ, Challma S, Lindholm T *et al*. Body odour disgust sensitivity is associated with xenophobia: evidence from nine countries across five continents. *R Soc Open Sci* 2023;**10**:221407.
49. Liuzza MT, Lindholm T, Hawley CB *et al*. Body odour disgust sensitivity predicts authoritarian attitudes. *R Soc Open Sci* 2018;**5**:171091.
50. Peer E, Rothschild D, Gordon A *et al*. Data quality of platforms and panels for online behavioral research. *Behav Res Methods* 2022;**54**:1643–62.

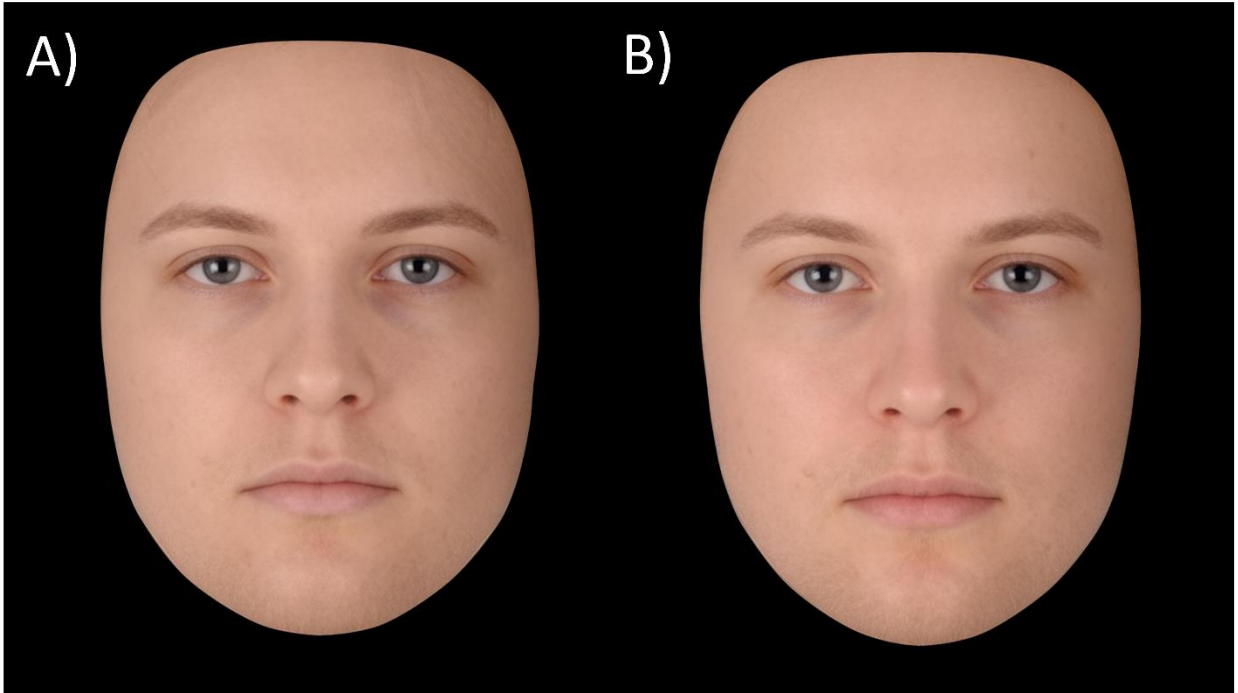
Figure 1. Facial composites displaying the average shape, color, and texture of 20 donors (eight women) photographed twice in a cross-over design, during experimentally induced (A) acute sickness and (B) placebo.

Figure 2. Women's and men's percentage of correct identifications (A) and discriminability  $d'$  (B). Medians (thick lines), means (large black dot), first and third quartiles, whiskers extending to the largest and the smallest values no further than 1.5 times the inter-quartile range, and individual data points (small dots) are indicated. Grey dashed lines depict random choice.

Figure 3. Posterior distributions for estimated drift rates ( $v$ ) for men (turquoise) and women (red) in the sick (dashed lines) and healthy (solid lines) conditions. The peak of each distribution indicates most probable drift rate value according to our model. Higher drift rate values reflect more information accumulated per unit of time so that less time is needed to make a correct response.

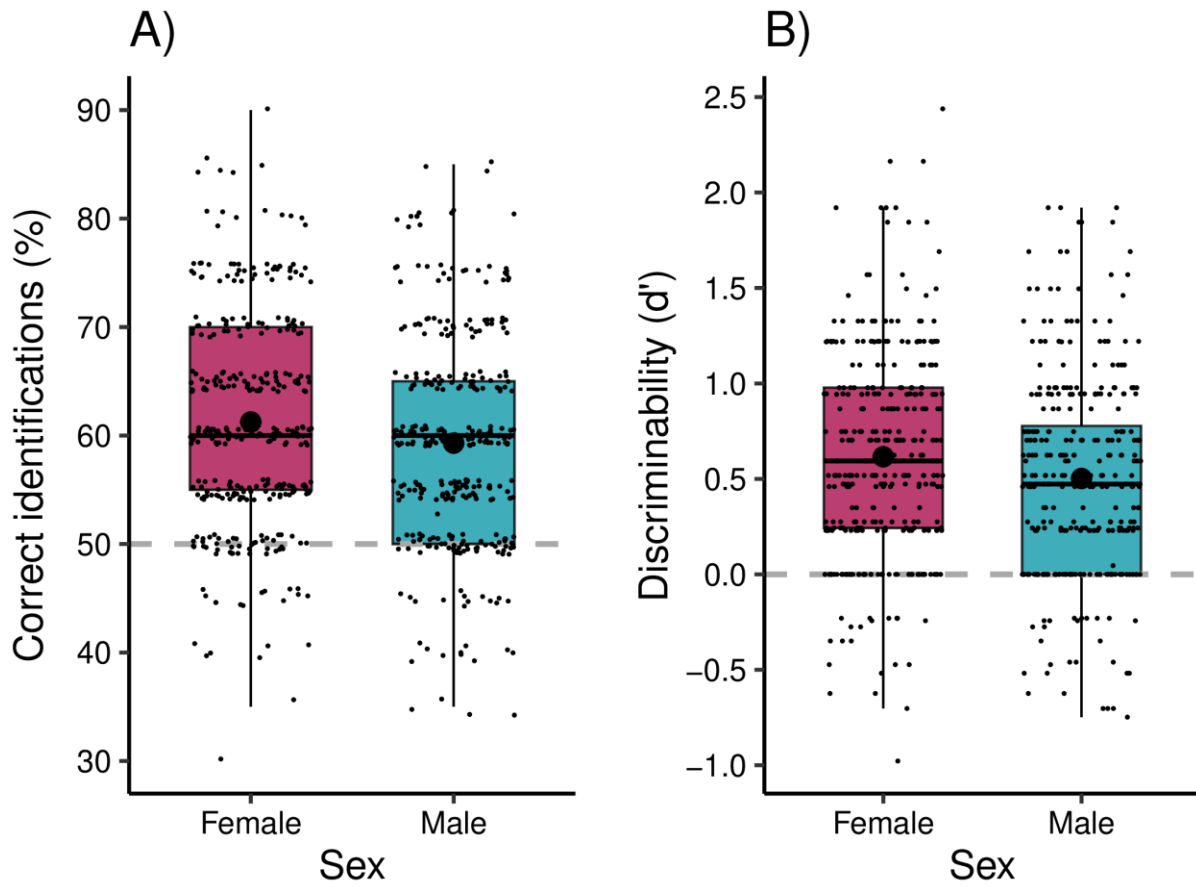
Figure 4. Scatterplots showing the relationship between the decision criterion  $c$  and the score on the Body Odor Disgust Scale (BODS). The decision criterion reflects how biased a participant is from a liberal ( $c < 0$ , most of the faces are considered sick) to a conservative bias ( $c > 0$ , few faces are considered sick). Linear regressions showed that individuals who are more easily disgusted towards body odors (A), especially towards *others'* body odors (B), are more sensitive to potential sickness cues (i.e., have a lower decision criterion  $c$ ). The decision criterion showed no statistically significant relationship with the internal body odor source subscale (C). Dots represent individual datapoints. The density of the data is illustrated by the intensity of the gray level of each data point.

Figure 1



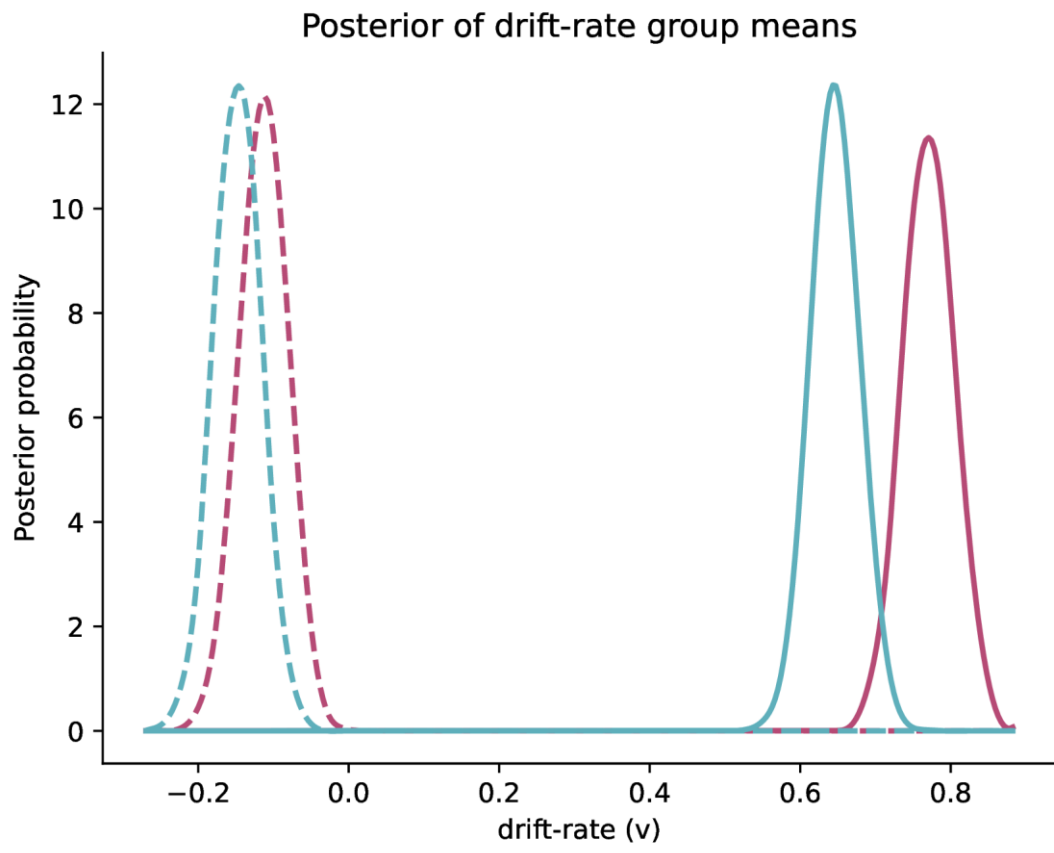
Accepted Manuscript

Figure 2



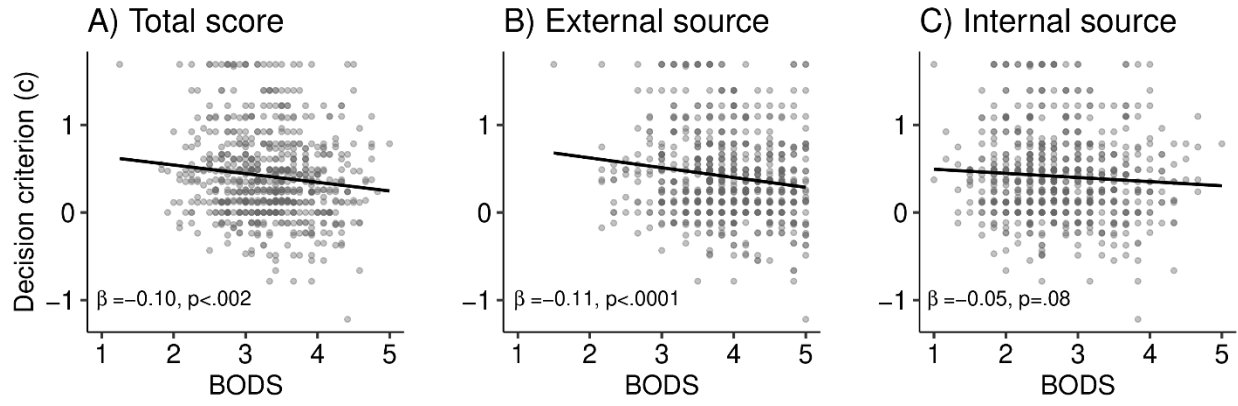
Accepted

Figure 3



Accepte

Figure 4



Accepted Manuscript