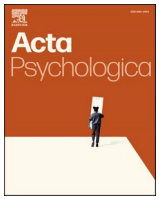
Acta Psychologica 235 (2023) 103893

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy



Should I trust you? Investigating trustworthiness judgements of painful facial expressions

Mathias Van der Biest [a], Emiel Cracco [b], Paolo Riva [c], Elia Valentini [d,*]

[a] Department of Experimental Psychology, University of Ghent, Faculty of Psychology and Educational Sciences, Ghent, Belgium
[b] Department of Experimental Clinical and Health Psychology, University of Ghent, Ghent, Belgium
[c] Department of Psychology, University of Milano–Bicocca, Milan, Italy
[d] Department of Psychology and Centre for Brain Science, University of Essex, Colchester, United Kingdom

ARTICLE INFO

Keywords:
Trustworthiness
Pain perception
Mouse tracking

ABSTRACT

Past research indicates that patients' reports of pain are often met with skepticism and that observers tend to underestimate patients' pain. The mechanisms behind these biases are not yet fully understood. One relevant domain of inquiry is the interaction between the emotional valence of a stranger's expression and the onlooker's trustworthiness judgment. The emotion overgeneralization hypothesis posits that when facial cues of valence are clear, individuals displaying negative expressions (e.g., disgust) are perceived as less trustworthy than those showing positive facial expressions (e.g., happiness). Accordingly, we hypothesized that facial expressions of pain (like disgust) would be judged more untrustworthy than facial expressions of happiness. In two separate studies, we measured trustworthiness judgments of four different facial expressions (i.e., neutral, happiness, pain, and disgust), displayed by both computer-generated and real faces, via both explicit self-reported ratings (Study 1) and implicit motor trajectories in a trustworthiness categorization task (Study 2). Ratings and categorization findings partly support our hypotheses. Our results reveal for the first time that when judging strangers' facial expressions, both negative expressions were perceived as more untrustworthy than happy expressions. They also indicate that facial expressions of pain are perceived as untrustworthy as disgust expressions, at least for computer-generated faces. These findings are relevant to the clinical setting because they highlight how overgeneralization of emotional facial expressions may subtend an early perceptual bias exerted by the patient's emotional facial cues onto the clinician's cognitive appraisal process.

1. Introduction

Research on person perception has shown that people evaluate faces on multiple trait dimensions, and these evaluations affect how we ultimately judge others (Oosterhof & Todorov, 2008). For example, facial appearance might influence personality attribution (Sutherland et al., 2015), whether we judge someone as a criminal (Eberhardt et al., 2006), and if we will vote for a politician (Todorov et al., 2005). More generally, we seem to quickly form an impression of strangers by assigning traits to them that resemble their emotional expression, a process labelled as *emotion overgeneralization* (Zebrowitz & Montepare, 2008). A particularly relevant trait in this respect is trustworthiness. Indeed, Oosterhof and Todorov (2008) showed that when facial cues of valence are clear, individuals displaying negative expressions (e.g., anger) are perceived as less trustworthy than those displaying positive facial expressions (e.g., happiness; see also Franklin & Zebrowitz, 2013; Oosterhof & Todorov, 2009; Todorov & Duchaine, 2008). This is important because trustworthiness judgments in everyday life inform people's perception of strangers' intention to help or harm (Todorov et al., 2015 for a review).

1.1. Trustworthiness judgments of painful facial expressions

We adopted a situational perspective in our experimental assessment of trust; that is, we conceived trust as an expectancy state generated by specific visual cues in a stranger's face, rather than a dispositional trait (Rotter, 1971). Yet, because our experimental tasks and procedures entailed no engagement in a relationship with the stranger's models, this type of trust is more pertinent to how much the onlooker would trust a stranger's face in general ('generalised trust' – Couch & Jones, 1997;

* Corresponding author at: Centre for Brain Science, Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom.
E-mail address: evalent@essex.ac.uk (E. Valentini).

https://doi.org/10.1016/j.actpsy.2023.103893
Received 19 September 2022; Received in revised form 14 March 2023; Accepted 15 March 2023

Freitag & Traunmüller, 2009) than to the trust implied in, for example, a transactional relationship ('interpersonal trust', Johnson-George & Swap, 1982).

The overgeneralization of facial expressions to trustworthiness has been suggested to be specific to expressions of anger and happiness (Engell et al., 2010). However, such overgeneralization could be particularly relevant in the clinical pain context. Indeed, the pain management literature indicates that pain reported by patients is often met with doubt and skepticism on the part of observers (Blomqvist & Edberg, 2002; Clarke & Iphofen, 2005; Montali et al., 2011), which might lead to an underestimation of the patient's pain (Riva, Rusconi, et al., 2011; Rusconi et al., 2010), especially when displayed by the elderly and women (Blomqvist & Edberg, 2002; Riva, Sacchi, et al., 2011), or ethnic minorities like black individuals (Banaji et al., 2021; Hoffman et al., 2016). In this context, different trust construal features will likely be at stake, ranging from a generalised trust to a more specific trust developed within the patient-carer relationship. Relatedly, behavioral manifestations of pain are sometimes expressions of malingering, catastrophizing, and somatization (Katz et al., 2015; Tuck et al., 2019).

These phenomena highlight the purported relevance of the authenticity of facial expressions of pain in determining the outcome of the onlooker's perception and decision-making (Williams, 2002). Interestingly, research indicates that onlookers struggle to distinguish between posed and genuine expressions of pain (Mende-Siedlecki et al., 2020 for a critical assessment) and judge posed faces as expressing more intense pain (Fernandes-Magalhaes et al., 2022). Whether spontaneous or acted pain expressions, caregivers and clinicians' judgments of facial expressions are an element that can have serious pain management and treatment implications (Wells et al., 2008), as their assessment of a patient's condition can be jeopardized by disbelief, lack of empathy, and trust in the patient's pain expressions (De Ruddere et al., 2012, 2014). Surprisingly, however, despite the role of disbelief and distrust in contributing to the stigma surrounding chronic pain (e.g., De Ruddere & Craig, 2016; Sims et al., 2021; Wakefield et al., 2021), research into how onlookers judge the trustworthiness of facial expressions of pain is still minimal. Here, we investigate this question by comparing trustworthiness judgments of painful expressions with trustworthiness judgments of positive expressions and other similar negative expressions, such as disgust.

The facial movements exerted during pain (i.e., wrinkling of the nose, closing of the eyes, rising of the cheeks, and lowering of the eyebrows, Patrick et al., 1986; Prkachin, 1992) are similar to those displayed when experiencing disgust, as identified by the Facial Action Coding System (FACS; Ekman & Friesen, 1978). However, despite this similarity, both the subjective experience and threat signaled by disgust and pain expressions can be distinguished. In fact, Kunz et al. (2013) found that participants responding to noxious stimuli with expressions of pain were mostly displaying contraction of the eyes and eyebrows whilst expressions in response to disgusting pictures were characterised by a curl of the upper lip and raise of the eyebrows. We reasoned that disgust expressions are an optimal comparison to investigate how pain expressions are evaluated concerning other negative expressions in terms of trustworthiness. Given that expressions of disgust have been shown to elicit lower trustworthiness ratings compared with happy expressions (e.g., Kugler et al., 2020; Ueda et al., 2017), we predicted a similar effect for pain expressions.

1.2. The present study

We investigated whether facial expressions of pain are perceived as untrustworthy compared to other facial expressions (e.g., neutral, happy). To this end, we measured trustworthiness judgments of four different facial expressions (i.e., neutral, happiness, pain, and disgust) in two studies, via both explicit self-reported ratings (Study 1) and implicit motor trajectories in a trustworthiness categorization task (Study 2). The latter task was included because a critical feature of trustworthiness judgments is that they are formed rapidly (Willis & Todorov, 2006). As a result, previous studies have suggested that measures tapping into implicit aspects of decision-making might be captured by computer mouse trajectories measures better than explicit ratings (e.g., Maldonado et al., 2019; Zgonnikov et al., 2017).

We further included both computer-generated and real faces in our studies to increase the findings' external and internal validity. Computer-generated faces allow for greater internal validity because researchers can vary the strength with which they express emotions. However, real faces allow for greater external validity because they are more like faces participants may meet in their daily lives. The importance of including both types of faces is further supported by evidence that neural and perceptual responses to computer-generated and real faces are not fully overlapping (e.g., Kätsyri et al., 2020). In particular, computer-generated faces seem to elicit overall lower trustworthiness ratings than real faces (Balas & Pacella, 2017).

1.2.1. Hypotheses and expectations

In light of the established link between positive/negative facial expressions and perceived trustworthiness (e.g., Oosterhof & Todorov, 2008), we first expected lower trustworthiness ratings in Study 1 for both disgust and pain expressions compared with happy expressions (H1). By contrast, we expected no significant difference in perceived trustworthiness between disgust and pain (H2). Similarly, in the trustworthiness categorization task (Study 2), we hypothesized that motor trajectories for trust categorizations would be more common, faster, and clearer (i.e., nearer to a straight trajectory) for happy than negative expressions. In contrast, motor trajectories for distrust categorizations would be more common, faster, and clearer for negative than for happy expressions (H3), with again no difference between pain and disgust expressions (H4).

1.2.2. Control analyses

To ensure that participants could accurately distinguish pain and disgust expressions, we also included a Specific Emotion task. In this task, we expected similar accuracy in categorizing pain and disgust expressions (H5), thus ruling out idiosyncratic identification (i.e., mismatch) of these two categories for both real and computer-generated faces. Moreover, for both Study 1 and Study 2 real faces datasets we also added the factor sex of the displayed faces to account for this potential interaction with the participants' sex and the emotional expression (non-directional two-tailed hypothesis). As additional analysis we excluded the 50 % expression intensity category for computer-generated faces as to address a potential bias of ambiguous expressions on rating and mouse tracking performance (non-directional two-tailed hypothesis).

2. Study one – web survey

2.1. Methods

2.1.1. Participants

Two hundred eighty respondents aged 18–65 (mean age = 23.48, SD = 9.93) participated in one of the two web-based cross-sectional surveys (developed using Qualtrics, Provo, UT) assessing real faces (224 females and 56 males). The distribution of ethnicity was as follows: 32 Asian/Pacific islanders, 23 Black or African Americans, 7 Hispanic or Latinos, 199 White Caucasians, and 19 from other ethnic groups. Similarly, 225 respondents aged 18–65 (mean age 27.99, SD = 12.62) participated in the second survey assessing computer-generated faces (151 females, 74 males). The ethnicity in this sample was distributed as follows: 18 Asian/Pacific Islanders, 19 Black or African Americans, 5 Hispanic or Latino, 170 White Caucasians, and 13 from other ethnic groups. They were recruited through a mix of social media ads and the Department SONA platform. Respondents were informed that the survey would take approximately 15 min and gave their informed consent before beginning

the study, which was approved by the University of Essex ethics committee (project code EV1501).

2.1.2. Stimuli

We used a total of 45 faces, 24 real faces (i.e., emotion; 6 neutral, 6 disgust, 6 pain, and 6 happiness expressions), and 21 computer-generated androgynes faces (3 neutral, 6 disgust, 6 pain, and 6 happiness expressions). The real faces were selected from Simon et al. (2008), then grey-scaled and oval-shape edited. Thus, background and hair were excluded to limit the information to the facial expression only. For every facial expression category, 3 of each emotional expression were displayed by males and 3 by females (i.e. sex). The computer-generated faces were created with FaceGen 3.1 and were extracted from a dataset already used in previous work (Riva, Sacchi, et al., 2011). They were all androgynous, that is, they were meant to equally express male and female visual features.[1] The emotional expressions varied along the intensity dimension, each representing one degree of the expression per each emotion category (50 %, 60 %, 70 %, 80 %, 90 %, 100 %). We added different degrees of expression (i.e., intensity) to increase the variability of the stimulus set and reduce habituation to these artificial faces. The distinctions between levels of emotional intensity were tested in Riva, Rusconi, et al. (2011) and Riva, Sacchi, et al. (2011). However, we did not include this variable as an independent predictor.

2.1.3. Design, task, and procedure

The cross-sectional web-based survey required an average completion time of about 20 min, but there was no restriction on the time allowed to complete the survey. All respondents were required to answer all questions but free to exit the survey anytime. We first asked respondents to provide us with their informed consent and some demographic information. Respondents were then presented with 21 more blocks (in randomized order) of questions concerning the different facial expressions. Respondents provided 5-point Likert scale ratings ranging from "negative" to "positive", "calming" to "arousing", "unattractive" to "attractive", "threatening" to "safe", "untrustworthy" to "trustworthy" (H1, H2). In addition, for a better descriptive characterization of the valence dimension, we asked respondents to indicate how much of the different basic emotions the specific face they were looking at was expressing, ranging from "not at all" to "extremely". These questions were delivered in a randomized order with each face stimulus. Note that all but trustworthiness ratings were added for exploratory purposes and will not be discussed (see supplementary material). The participants did not receive a definition of trustworthiness. This choice was in keeping with the person perception literature (e.g., Engell et al., 2010; Oosterhof & Todorov, 2008; Todorov et al., 2008). This approach seems to be also adopted in studies targeting interpersonal trust (Hale et al., 2018). All the relevant material associated with the study can be found on the open repository (OSF).

2.1.4. Data analysis

For each questionnaire, we transformed the trustworthiness ratings (i.e., untrustworthy = 1, mildly untrustworthy = 2, neither untrustworthy nor trustworthy = 3, mildly trustworthy = 4, trustworthy = 5). Next, we constructed a linear mixed model (LME) for each questionnaire (i.e., real faces; computer-generated faces), with emotion and sex (i.e. only for the real faces) as a fixed effects. To determine the random effects structure of each model, we applied a backwards selection procedure (Matuschek et al., 2017), as this has been shown to balance the false-positive and false-negative rates. For real faces, this resulted in a random-effects model with gender and subject (i.e., gender|subject). For computer-generated faces this was the simplest random effect model (i.e., 1|subject). *P*-values were calculated using the Satterthwaite's method. All models and corresponding *p*-values were constructed using the "lmertest" package (Kuznetsova et al., 2017) in R (R Core Team, 2021). The corresponding effect sizes and 95 % confidence intervals were calculated with the "effect size" package (Ben-Shachar et al., 2020). Post hoc z tests *p* values were corrected with the false discovery rate method for multiple comparisons as provided by the emmeans package. Statistical significance was set at $p < .05$.

2.1.5. Results

We report the descriptive statistics of the valence, arousal, and attractiveness of each emotional expression in the supplementary materials (Appendix A, Figs. S1-S2).

For both questionnaires, we removed participants who did not complete all the questions. This resulted in a sample size of 190 participants for the questionnaire assessing real faces and 134 for computer-generated faces.

2.1.5.1. Real faces. The LME model indicated a significant main effect of emotion, $F(3,4174) = 323.58$, $p < .0001$, $\eta_p^2 = 0.19$, CI95% = [0.13,0.21]. Trustworthiness ratings were highest for happy, followed by neutral, pain, and disgust expressions (Fig. 1A). Concerning H1, the post hoc test showed a significant difference between happy and disgust ($z = 28.64$, $p < .0001$) and pain expressions ($z = 23.07$, $p < .0001$). Concerning H2, we found a significant difference between disgust and pain ($z = -5.57$, $p < .0001$), indicating that trustworthiness ratings were significantly lower for disgust compared with pain expressions. All remaining post hoc tests indicated a significant difference in trustworthiness ratings ($z > -5.74$, $p < .0001$), except for the difference between neutral and pain ($z = 0.17$, $p = .863$). There was a main effect of sex, $F(1,189) = 77.66$, $p < .0001$, $\eta_p^2 = 0.29$, CI95% = [0.19,0.39]. Trustworthiness ratings were higher for female than male faces. Lastly, there was a significant interaction effect between sex and emotion, $F(3,4174) = 12.79$, $p < .0001$, $\eta_p^2 = 0.0$, CI95% = [0.0,0.02]. Post hoc tests revealed a significantly higher trustworthiness rating for female neutral ($z = -9.31$, $p < .0001$), happy ($z = -3.01$, $p = .0035$), and pain expressions ($z = -4.57$, $p < .0001$), compared to males. There was no significant difference for disgust expressions ($z = -0.94$, $p = .35$ - see Appendix B, Table S1-S2).

2.1.5.2. Computer-generated faces. The LME model indicated a significant main effect of emotion $F(3,2677) = 906.57$, $p < .0001$, $\eta_p^2 = 0.50$, CI95% = [0.48,53]. Trustworthiness ratings were highest for happy faces, followed by neutral, pain, and disgust (Fig. 1B). Concerning H1, post hoc tests revealed a significant difference between trustworthiness ratings for happy and disgust ($z = 43.15$, $p < .0001$) and pain expressions ($z = 43.00$, $p < .0001$). Concerning H2, there was no significant difference between disgust and pain ($z = -0.15$, $p = .999$). All other post hoc tests were significant ($z > -9.63$, $p < .0001$, see Appendix B, Table S3, for all contrasts).

The 50 % intensity control analysis indicated that the exclusion of the 50 % expression data from the LME model produced a significant main effect of emotion $F(3,133) = 136.74$, $p < .0001$, $\eta_p^2 = 0.74$, CI95% = [0.69,1.00]. Trustworthiness ratings were highest for happy faces, followed by neutral, pain, and disgust. Concerning H1, post hoc tests revealed a significant difference between trustworthiness ratings for happy and disgust ($z = 19.45$, $p < .0001$) and pain expressions ($z = 19.00$, $p < .0001$). Concerning H2, there was no significant difference between disgust and pain ($z = -1.30$, $p = .195$). All other post hoc tests were significant ($z > -8.68$, $p < .0001$), thus indicating that the exclusion of these data did not alter the results pattern (see also Appendix B, Table S4, for all contrasts).

[1] Note that despite these stimuli morphed together masculine and feminine faces in equal parts, they were not perceived as fully androgynous in the original study. For example, facial expressions of pain were more likely to be categorized as male than female (Riva, Rusconi, et al., 2011; Riva, Sacchi, et al., 2011).

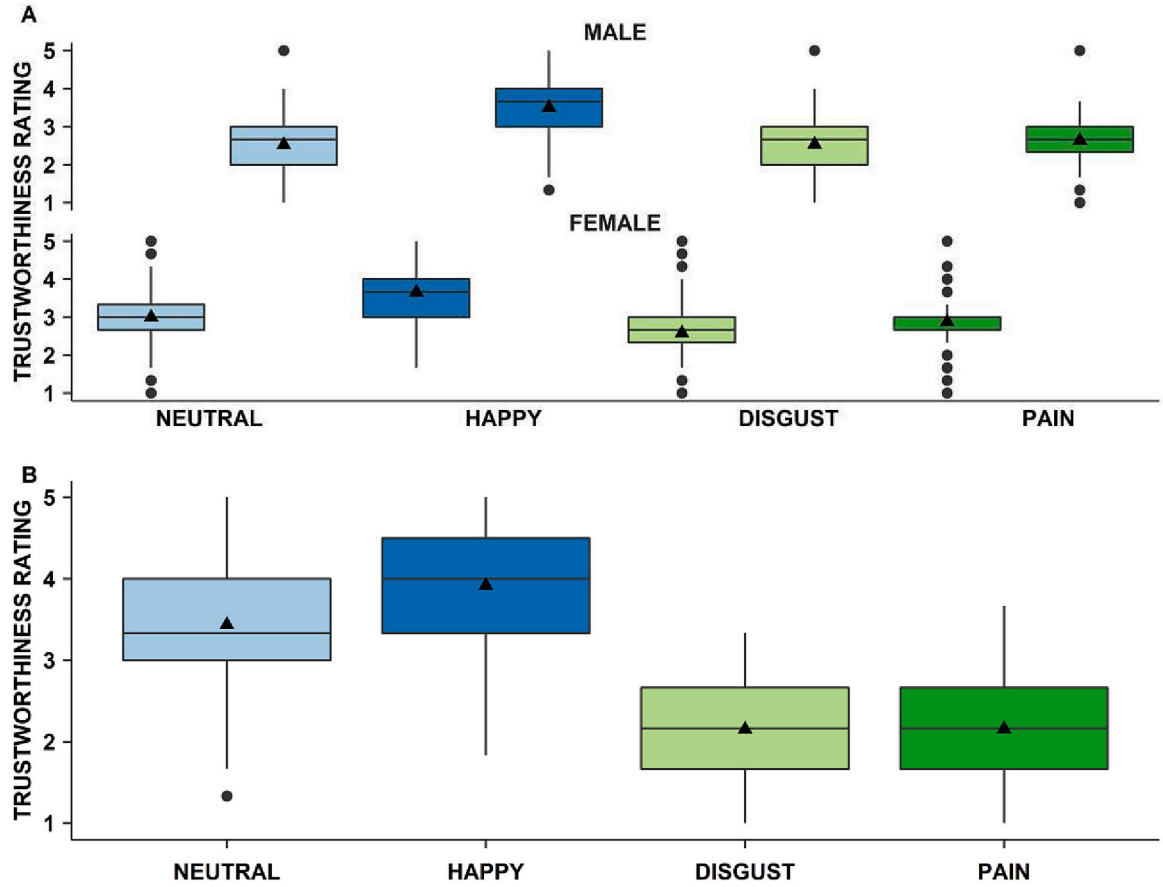


Fig. 1. Results of the trustworthiness survey (Study 1). Trustworthiness ratings for each emotional expression during real (A) and computer-generated faces (B) on a Likert scale from 1 (untrustworthy) to 5 (trustworthy).
Note. Study 1 trustworthiness ratings (y axis) for all emotional expressions (neutral – light blue, happy – dark blue, disgust – light green, pain – dark green, x axis) and the two types of stimuli (real and computer-generated faces – A and B panel respectively). Panel A is futher divided into two smaller panels, representing the sex of the face stimulus. The upper panel represents male while the lower panel female faces. Note the greater trustworthiness assigned to happy facial expressions, especially during the observation of real faces. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Study 2 – mouse tracking task

3.1. Methods

3.1.1. Participants

Nineteen right-handed female and ten male psychology students aged 20–29 (mean = 22.93, SD = 1.85) were recruited through the Sona system at the Department of Psychology at the University of Milano – Bicocca. We did not collect ethnicity information from this sample. Also, due to technical problems with the computer-generated stimulus task, one data set was missing, resulting in a sample size of 28 for the analyses of the tasks with computer-generated faces. Respondents were recruited through the Department SONA platform and were informed that the study would have lasted about 30 min. They gave their informed consent before beginning the study, which was approved by the University of Milan ethics committee (project code RM-2021-489).

3.1.2. Design, task, and procedure

We used a total of 34 faces, 24 real faces, and 10 computer-generated androgynes faces. The real faces were composed of 6 neutral, 6 disgust, 6 pain and 6 happiness expressions (i.e. emotion). For every category, 3 of each emotional expression were displayed by males and 3 by females (i.e. sex). We used three androgynous faces for each emotion (i.e., pain, happy, disgust) and used only 1 neutral face (extracted from Riva, Sacchi, et al., 2011).

Participants completed a set of three separate categorization tasks, in which they categorized target stimuli along Trustworthiness (Trustworthy vs Untrustworthy), General Emotion (Happy vs Disgust vs Pain), or Specific Emotion (Disgust vs Pain). The last two were devised as a control task to ensure the participants' trajectories would suggest correct differentiation of the emotional expressions, especially between expressions of pain and disgust. This, in turn, would ensure the interpretability of the results obtained in the Trustworthiness task. Faces independently varied in terms of emotional expression (four levels: Neutral, Happy, Pain, and Disgust). The three tasks were delivered in separate sessions in a counterbalanced order across participants. Due to a failure in the pseudorandomization protocol of the General Emotion Task trials, we decided not to analyse these data, thus leaving us with the analysis of the Trustworthiness (Trustworthy vs Untrustworthy) and Specific Emotion (Disgust vs Pain) tasks. These were anyway the necessary tasks to test our hypotheses hence we report the results obtained with these two tasks only.

In a typical two-choice categorization task, participants are presented with an image and at least two response options at the top left or right corners. Participants are then instructed to move the mouse cursor from the image location to the appropriate option, click on it, and repeat the action across numerous trials. The x-, y-coordinates of the mouse pointer are recorded during the process, allowing the experimenter to assess if the participant selected the correct response category and how much they deviated from a perfect trajectory (i.e., straight line from the

start to the response location). Direct trajectories indicate certainty, whereas indirect trajectories highlight attraction to the unselected choice and, therefore, uncertainty (Freeman et al., 2013). Thus, the advantage of analyzing motion trajectories is that it provides the researcher with a direct measure of decisional uncertainty.

All stimuli were presented on an HP Compaq 6200 Pro Small factor computer with a 19-inch screen and 1920 × 1080 resolution and using Mousetracker software (Freeman & Ambady, 2010; http://www.freemanlab.net/mousetracker). Mouse sensitivity was set to the 6th notch to ensure 100 %-point speed accuracy. Response buttons were positioned in the upper left and upper right corners of the screen, displayed with bold letters in Arial (size 22) with a black background and measuring 0.4 × 0.2 cm.

Participants were tested alone in a lab cubicle and seated at an average distance of 60 cm from the computer screen, with the computer mouse placed to their right side. They were instructed to fixate the center of the screen where the "start" button would appear and either judge the type of emotional expression (i.e. Specific Emotion: Disgust vs Pain) or its trustworthiness (i.e. Trustworthy vs Untrustworthy). Before beginning the experiment, participants completed a practice shape categorization task (vegetables vs fruit) to familiarize themselves with the task procedure (40 trials). The procedure required participants to click on a 'Start' button located at the bottom-center of the screen to begin each trial. A face stimulus immediately replaced this. Participants were asked to move the mouse as quickly and accurately as possible towards one of the two response boxes (for example, Disgust vs Pain) located at the top-left and top-right corners of the screen and click on the chosen category. The spatial location of the two labels was counterbalanced across participants. Face stimuli were presented in a randomized order.

For the Specific Emotion task with real faces, participants were presented with a total of 120 facial expressions, 60 of each sex (i.e. male, female), with each actor/actress, repeated 10 times. For the Specific Emotion task with computer-generated faces, participants were presented with 78 androgynous faces displaying a 75 % emotional intensity (i.e. 39 for pain and 39 for disgust expressions). For the trustworthiness task with real faces, participants were presented with 240 faces, 6 for each emotional expression (3 for each sex), repeated 10 times. For the trustworthiness task with computer-generated faces, participants were presented with 240 androgynous faces displaying different emotions (pain, disgust, happiness expression) and intensities (50 %, 75 %, and 100 %). Each emotion and intensity were repeated 20 times, while neutral faces were presented 60 times. As in Study 1, we did not provide the participants with a definition of trustworthiness.

Trials were 3 s in duration and were separated by variable inter-trial intervals (2–4 s). A message encouraging quicker categorization appeared when movements were not initiated within 400 ms. A fixation cross replaced the face stimulus after any response, or if participants did not respond on time, remaining on the screen until the beginning of the subsequent trial. Participants were required to return the mouse on the 'Start' button and click on it to start the subsequent trial. Each categorization task lasted approximately 10 min, depending on how quick or slow the participants were in responding. During the task, the mouse's streaming x, y coordinates were recorded to allow for point-by-point trajectory estimation.

3.1.3. Data analysis

3.1.3.1. Data preparation. For each mouse-tracking task, we collected the participant's response categorization (i.e. Disgust vs Pain, Trustworthy vs Untrustworthy), reaction times (RTs), and two measurements of decision uncertainty, namely the maximum deviation (i.e., largest deviation from the ideal trajectory, MD), and area under the curve (i.e., area under the actual trajectory and the ideal trajectory, AUC) (Freeman & Ambady, 2010).

Participants with a mean above or below 3 standard deviations (SD) for AUC, MD, or RT, were removed from the analyses of the Trustworthiness task. Participants with an accuracy below 60 % of the Specific emotion (i.e. Disgust vs Pain) task were also excluded from further analyses. The exclusion criteria at the trial level were based on previous studies and mouse tracking guidelines (e.g., Hehman et al., 2015; Kieslich & Henninger, 2017). For all analyses, we excluded trials with a response time slower than 2500 ms or an initiation time 3 SD above or below the participants' mean.

Six participants were excluded for the analyses of real faces in the Specific Emotion task due to numerous incorrect responses (>40 %). As a result of our filtering procedure, 6 % of trials were removed. Please note that this applies only to the categorization data analyses (see methods for details). For the analyses of the computer-generated faces, four participants were considered outliers due to poor performance (i.e., accuracy <60 %).

For the analyses of real faces in the Trustworthiness task, none of the participants were excluded. And in total, 6 % of all trials were considered as outliers (3SD from the participants mean for the initiation time, and AUC, MD or RT). One participant was excluded from all analyses of computer-generated faces due to the exclusion criteria of the MD and AUC. Furthermore, 4 % of the trials were considered as outliers.

3.1.3.2. Statistical analyses. The reaction times and trajectory measures (i.e., AUC, MD) were investigated with linear mixed-effects models. The categorization choice data (i.e., accuracy) were analyzed with generalised linear mixed-effects models using a binomial logit link function. The GLME models were constructed with "lme4" (Baayen et al., 2008; Bates et al., 2015). We created different factors according to the type of task. The Specific Emotion task included the type of emotional expression (i.e., Disgust and Pain). The trustworthiness task included both the type of emotional expression (i.e., Neutral, Happy, Pain, Disgust) and behavioral choice (i.e., Trustworthy, Untrustworthy). Only for the real faces, we included sex as an additional fixed effect (i.e. Male, Female). We applied the same criteria as the first study to determine each model's random effect structure. For the LME models, we calculated the *p*-values with the Satterthwaite's method. Moreover, we computed the corresponding effect sizes and 95 % confidence intervals. For the GLME models, we used the Wald test. Post hoc z tests *p*-values were corrected with the false discovery rate method for multiple comparisons as provided by the emmeans package. Lastly, we calculated the accuracy for each emotion to investigate if the Disgust and Pain categorization was above the chance level (50 %) using one-sample *t*-test. Additionally, we calculated the Bayes factor for each test with JASP (Version 0.14.1.0, JASP team, 2021).

3.1.4. Results

3.1.4.1. Specific Emotion task. The pain vs disgust categorization task tested whether participants could distinguish between pain and disgust expressions above chance level. Results confirmed our expectation of similar accuracy in categorizing pain and disgust expressions (H5 – see supplementary materials Appendix C, Fig. S3).

3.1.4.2. Trustworthiness task. To investigate the difference in trustworthiness perception for positive and negative emotions, we developed a trustworthiness categorization task with four types of emotions (i.e., neutral, happy, disgust, and pain).

3.1.4.3. Real faces

3.1.4.3.1. Categorization. The probability of categorizing a face as trustworthy was 92 %, CI95% = [85,96], for happy expressions, and 54 % CI95% = [32, 74] for neutral expressions. The probability of categorizing a face as untrustworthy was 94 %, CI95% = [88,99] for disgust expressions, and 91 %, CI95% = [77,97] for pain expressions.

The GLME analysis revealed a significant main effect of emotion (Fig. 2A), $X^2(3,29) = 51.42$, $p < .001$. Concerning H3, post hoc analysis indicated that there was a significant difference in trustworthiness perception between happy and disgust ($z = 6.61$, $p < .0001$), and pain expressions ($z = 6.52$, $p < .0001$). Concerning H4, there was no significant difference in trustworthiness perception between pain and disgust ($z = -1.51$, $p = .131$). All remaining comparisons were significant (all $z \geq 3.59$, all $p \leq .001$, see Fig. 2A and Appendix D Table S5, for all contrasts). There was a significant interaction effect of sex and emotion, $X^2(3,29) = 15.57$, $p = .001$. Post hoc analyses showed that female neutral faces were categorized as more trustworthy compared with male faces, $z = 3.15$, $p = .002$. All remaining contrasts were not significant (z

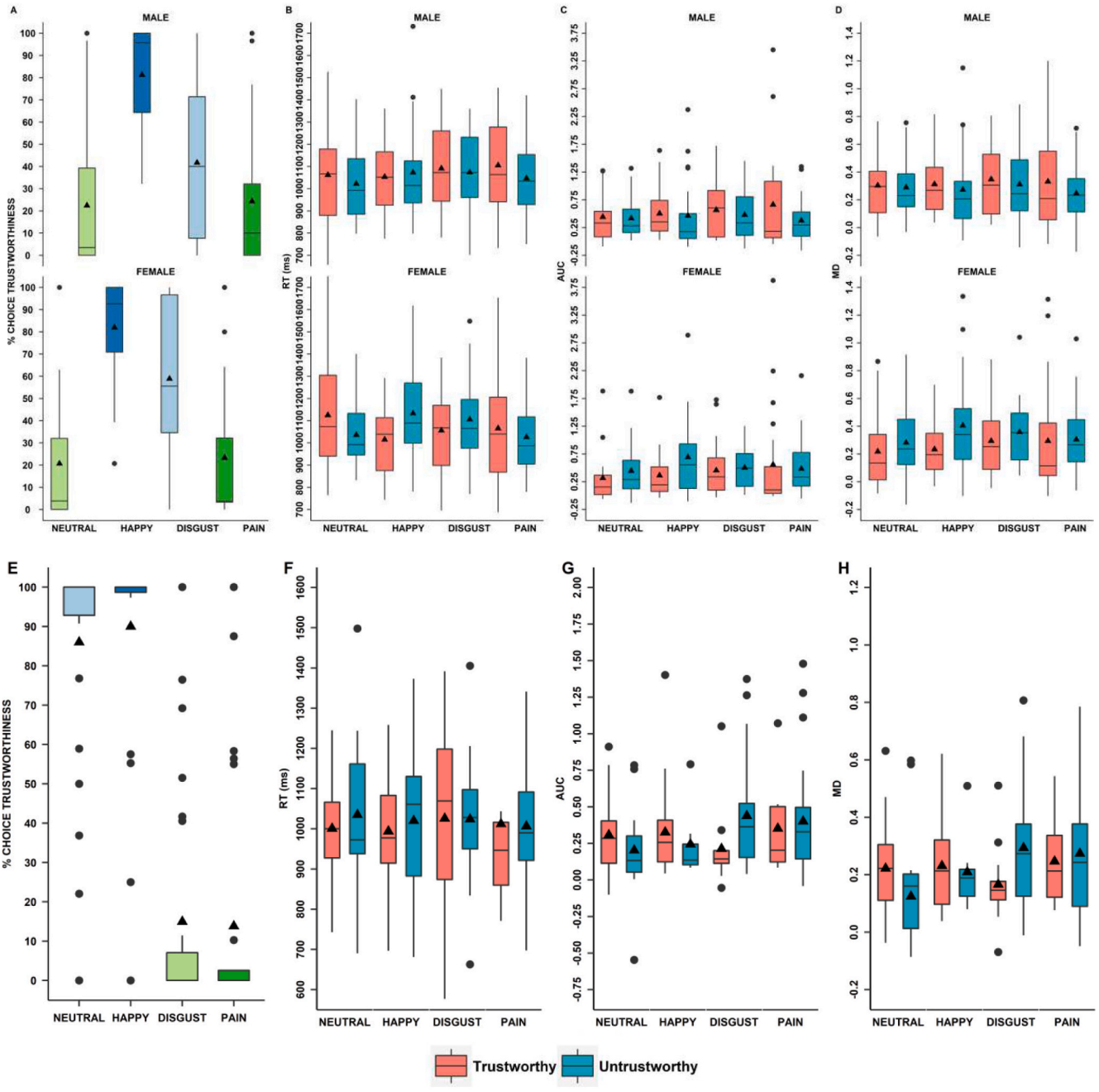


Fig. 2. Results of the mouse-tracking task (Study 2). Percentage of trustworthiness categorizations for each emotion (A, E), and changes in performance for reaction times (B, F), area under the curve (C, G), and maximum deviation (D, H).

Note. Trustworthiness mouse-tracking performance for each emotion. The upper four boxplots (i.e., A, B, C, D) are a graphical presentation of the trustworthiness task with real faces, while the lower four boxplots (i.e., E, F, G, H), are a graphical presentation of the results of the trustworthiness task with computer-generated faces for the four dependent measurements (i.e., categorization, reaction times, area under the curve, and motor deviation). For the real faces, the graphs are divided for sex: the upper graphs represent the male expressions, the lower graphs female. The red boxplots depict trustworthy categorizations, the blue boxplots depict untrustworthy categorizations. The horizontal black bar represents the median for each condition, and body of the boxplot represents the interquartile range (IQR), with the outer lines the 25 (Q1) and 75 (Q2) percentiles. The black dots are outliers that are above or below Q1/Q3 ± 1.5* IQR. The black triangles represent the mean for each emotion and level of trustworthiness. Plot A and E are the graphical representation of the trustworthiness categorization for each emotion. Plot B and F, C and G, and D and H, represent respectively the results of the RT, AUC, and MD. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

≤ 0.27, $p \leq .843$, see Appendix D Table S6, for all contrasts). We observed no significant main effect of sex, $X^2(1,29) = 1.39$, $p = .239$.

3.1.4.3.2. RT. We found no significant main effect of emotion, F (3,31.3) = 1.96, $p = .139$, $\eta_p^2 = 0.16$, CI95% = [0,0.36] (counter H3). The main effect of choice was not significant, F(1,22.8) = 0.38, $p = .544$, $\eta_p^2 = 0.02$, CI95% = [0,0.22]. Likewise, there was no significant main effect of sex F(1,61.3) = 1.27, $p = .263$, $\eta_p^2 = 0.02$, CI95% = [0,0.14]. However, we did find a significant emotion x choice interaction, F(3, 684.7) = 7.37, $p < .001$, $\eta_p^2 = 0.03$, CI95% = [0.01,0.06] (H3), indicating that participants were faster to categorize faces as trustworthy when they viewed happy expression (z = 3.31, $p = .004$). There was no difference in reaction times between trustworthiness for the neutral, disgust, nor pain expressions (counter H4) (all $z \leq 1.21$; $p \geq .355$; Fig. 2B and Appendix D Table S7, for all contrasts). Moreover, there was a significant interaction effect of choice and sex F(3,1328.8) = 6.11, $p = .014$, $\eta_p^2 = 0$, CI95% = [0.00,0.01]. Post hoc analyses, showed that female expressions were categorized faster as trustworthy compared to male, z = 2.43, $p = .031$. There was no difference in RTs when categorizing expressions as untrustworthy, z = 0.96, $p = .336$. There was no significant interaction effect of emotion and sex, F(3,5824.5) = 0.89,$p = .448$ $\eta_p^2 = 0$, CI95% = [0,0], nor a significant interaction effect of choice, emotion, and sex, F(3,2071.6) = 1.47, $p = .220$, $\eta_p^2 = 0.00$, CI95% = [0,0.01].

3.1.4.3.3. AUC. Analyses revealed no significant main effect of emotion; F(3,68.6) = 1.48, $p = .227$, $\eta_p^2 = 0.06$, CI95% = [0,0.17], choice, F(1,27.5) = 0.01, $p = .756$ $\eta_p^2 = 0.0$, CI95% = [0,0.15], nor sex, F (1,55.7) = 1.20, $p = .278$ $\eta_p^2 = 0.02$, CI95% = [0,0.14]. However, we found a significant interaction effect of choice and sex, F(1,3636.3) = 21.33, $p < .0001$, $\eta_p^2 = 0.00$, CI95% = [0,0.01]. The AUC was larger when categorizing a female expression as untrustworthy, z = 2.13, $p = .033$. While the AUC was larger when categorizing male expressions as trustworthy, z = 3.62,$p = .001$. There was no significant interaction effect of emotion and sex F(3,6232.3) = 0.46, $p = .713$, $\eta_p^2 = 0.00$, CI95% = [0,0.01], nor a significant interaction effect of choice and emotion F(3,1676.7) = 2.46,$p = .061$, $\eta_p^2 = 0.00$, CI95% = [0,0.01], (counter H3, H4). There was no significant interaction effect of emotion, choice and sex, F(3,5000.8) = 0.16, $p = .923$, $\eta_p^2 = 0.00$, CI95% = [0,0.0] (Fig. 2C).

3.1.4.3.4. MD. Analyses revealed no significant main effect of emotion F(3,65.8) = 2.01, $p = .121$, $\eta_p^2 = 0.08$, CI95% = [0,0.21] (counter H3 and H4), nor a main effect of choice, F(1,26.8) = 0.54, $p = .470$, $\eta_p^2 = 0.02$, CI95% = [0,0.21] nor a main effect of sex, F(1,96.9) = 1.13, $p = .255$, $\eta_p^2 = 0.01$, CI95% = [0,0.09]. We found a significant interaction effect of sex and choice, F(1,5796.4) = 19.91, $p < .0001$. Post hoc analyses revealed, the MD was larger when categorizing female expressions as untrustworthy compared with males (z = 2.23, $p = .026$). Likewise, the MD was larger for male expressions that were categorized as trustworthy (z = 3.69, $p = .001$). There was no significant interaction effect of emotion and choice, F(3,1648.2) = 2.03, $p = .108$, $\eta_p^2 = 0.00$, CI95% = [0,0.01] (counter H3, H4), emotion and sex F(3,6335.4) = 0.89, $p = .447$, $\eta_p^2 = 0.00$, CI95% = [0,0.00], nor a significant interaction effect of emotion, choice and sex, F(3,5747.3) = 0.41, $p = .744$, $\eta_p^2 = 0.00$ CI95% = [0,0.00] (Fig. 2D).

3.1.4.4. Computer-generated faces

3.1.4.4.1. Categorization. The probability of categorizing the expression as trustworthy was 100 % 95%CI = [99,100]% for happy expressions and 100 %, 95%CI = [96,100]% for neutral expressions. The probability of categorizing the pain expression as untrustworthy was 99 %, 95%CI = [94,100]% and 95 %, 95%CI = [86,98]% for disgust expressions. When excluding the 50 % expression intensity level the probability of categorizing the expression as trustworthy was 100 %, 95%CI = [99,100] for happy expressions and 99 %, 95%CI = [99,100]% for neutral expressions. The probability of categorizing the pain expression as untrustworthy was 100 %, 95CI,[97,100]%, and 99 %, 95CI = [94,100]% for disgust expressions.

The GLME analysis showed a significant main effect of emotion (Fig. 2E), $X^2(3,27) = 33.48$, $p < .001$. Concerning H3, post hoc comparisons revealed a significant difference between happy and disgust, z = 5.67, p < .001, and pain expressions, z = 5.59, $p < .001$. In contrast with H4, post-hoc analysis indicates that computer-generated pain expressions were perceived as less trustworthy than disgust expressions (z = 2.84, $p = .006$). All remaining post hoc comparisons were significant (all $z \geq 5.51$, $p \leq .001$), except for neutral vs happy ($z = -1.63$, $p = .104$) (see Appendix D, Table S8, for all contrasts). When excluding the 50 % expression intensity level the GLME analysis still reported a significant main effect of emotion, $X^2(3,27) = 28.30$, p < .001. The H3 was also confirmed by a significant difference between happy and disgust, z = 4.99, $p < .0001$, and pain expressions, z = 4.86, $p < .0001$. However, when testing H4, the difference between disgust and pain vanished (z = 1.50, $p = .134$)., and there was no difference in perceived trustworthiness between neutral and happy faces ($z = 2.01$, $p = .054$). All remaining post hoc comparisons were significant ($z \geq 4.75$, $p \leq .001$) (Appendix D, Table S9).

3.1.4.4.2. RT. We found no main effect of emotion F(3,4.98) =4.73, $p = .064$, $\eta_p^2 = 0.74$, CI95% = [0.0,0.90], nor a main effect of choice F (1,8.40) = 3.35, $p = .103$, $\eta_p^2 = 0.28$, CI95% = [0,0.65]. There was no significant interaction effect of emotion and choice F(3,5.60) = 2.18, $p = .198$, $\eta_p^2 = 0.54$, CI95% = [0,0.80] (Fig. 2F) (counter H3, H4). When excluding the 50 % expression intensity level the GLME analysis confirmed these results: no significant main effect of emotion F(3,2.0) = 1.32, $p = .458$, $\eta_p^2 = 0.66$, CI95% = [0.00,0.90], nor a significant main effect of choice, F(1,2.92) = 0.77, $p = .447$, $\eta_p^2 = 0.21$, CI95% = [0.00,0.74], nor a significant interaction effect, F(3,4.01) = 0.30, $p = .828$, $\eta_p^2 = 0.18$, CI95% = [0.00,0.56].

3.1.4.4.3. AUC. We found no significant main effect of emotion F (3,15.41) = 1.61, $p = .227$, $\eta_p^2 = 0.24$, CI95% = [0,0.50], nor a main effect of choice F(1,21.68) = 1.21, $p = .284$, $\eta_p^2 = 0.05$, CI95% = [0,0.30], nor a significant interaction effect of emotion and choice F (3,182.81) = 2.11, $p = .101$, $\eta_p^2 = 0.03$, CI95% = [0,0.09] (Fig. 2G) (counter H3 and 4). When excluding the 50 % expression intensity level the GLME analysis confirmed these results: no significant main effect of emotion, F(3,520.54) = 0.42, $p = .74$, $\eta_p^2 = 0.00$[0.00,0.01], nor a significant main effect of choice, F(1,25.01) = 1.08, $p = .308$, $\eta_p^2 = 0.04$ [0.00,0.27], nor a significant interaction effect, F(3,1638.75) = 1.22, $p = .301$, $\eta_p^2 = 0.0$[0.00,0.01].

3.1.4.4.4. MD. We found no significant main effect of emotion F (3,20.31) = 1.41, $p = .268$, $\eta_p^2 = 0.17$, CI95% = [0, 0.41], nor a main effect of choice F(1,25.08) = 1.19, $p = .285$, $\eta_p^2 = 0.05$, CI95% = [0,0.27], nor a significant interaction effect F(3,258.29) = 2.16 $p = .094$, $\eta_p^2 = 0.02$, CI95% = [0, 0.06] (Fig. 2H) (counter H3, H4). When excluding the 50 % expression intensity level the GLME analysis confirmed these results: no significant main effects of emotion, F (3,8.55) = 2.59,$p = .120$, $\eta_p^2 = 0.48$, CI95% = [0.00,0.73], nor a significant main effect of choice, F(1,16.82) = 1.36, $p = .259$, $\eta_p^2 = 0.08$, CI95% = [0.00,0.37], nor a significant interaction effect, F(3,12.12) = 2.48, $p = .111$, $\eta_p^2 = 0.38$, CI95% = [0.00,0.64].

4. General discussion

Research highlights how the assessment of a patient's condition can be jeopardized by disbelief, lack of empathy, and trust in the patient's pain expressions, thus highly contributing to the stigma surrounding chronic pain (De Ruddere et al., 2012, 2014; De Ruddere & Craig, 2016; Sims et al., 2021; Wakefield et al., 2021). Our main expectation was that trustworthiness ratings would be lower for both disgust and pain expressions compared with happy expressions (H1) and that disgust and pain expressions would be categorized more as untrustworthy relative to happy expressions (H3). By contrast, we expected no significant difference in self-report and categorization performance between disgust and pain (H2, H4). Importantly, categorization findings should not have

been accounted for by a poor (i.e., chance-level) identification of disgust and pain expressions (H5).

In line with our H1 and H3, not only facial expressions of disgust, but also painful expressions are judged as less trustworthy than happy expressions. It is noteworthy that Study 1 revealed no significant difference in trustworthiness ratings between pain and neutral expressions of real faces. Interestingly, research indicates that neutral expressions may be perceived as not so neutral (Albohn et al., 2019; Lee et al., 2008). In contrast with our H2, we found different trustworthiness ratings for pain and disgust expressions of real faces in Study 1. In addition, we found lower trustworthiness categorization for disgust than pain expressions for computer-generated faces in Study 2. Crucially, when excluding trials displaying ambiguous expressions (i.e., 50 % degree of expression), there was a substantial overlap in categorization performance.

It is worth noting that these differences cannot be accounted for by participants inability to distinguish between disgust and pain expressions. Indeed, the participants' performance in the Specific Emotion task in Study 2 suggests they could distinguish between pain and disgust (H5), thus excluding a perceptual confounder for both real and computer-generated faces (Appendix C, Fig. S3A-B). These findings support the notion that despite the similarity between the two expressions (Ekman & Friesen, 1978), both the affective experience and threat signaled by disgust and pain expressions are not confused (Kunz et al., 2013).

Such a mechanism may have important implications for clinical pain practice, as previous research suggests that perceived trustworthiness does not reflect the actual trustworthiness of our interaction partners (Rule et al., 2013). Indeed, past research indicated that patients' pain reports are often met with doubt and skepticism by observers (Blomqvist & Edberg, 2002; Clarke & Iphofen, 2005; Montali et al., 2011) and that this can lead to an underestimation of pain (Riva, Rusconi, et al., 2011; Rusconi et al., 2010).

In addition, and in contrast to the explicit choice data, none of the implicit psychomotor dependent variables (RT, AUC, MD) highlighted an interaction between facial expression and categorization (H3, H4). However, due to the strong trustworthiness categorization for happy expressions, and strong untrustworthiness categorization for pain and disgust, this lack of effect could be due to limited power. Indeed, we obtained a lower number of trustworthiness categorization trials for negative expressions, compared with positive expressions (mirrored by greater vs lower untrustworthiness trials). While this confirms our initial hypothesis (H1), it questions whether these mouse tracking measurements could be a sensitive behavioral index of trustworthiness.

Although our work is insufficient to provide a robust answer to whether disgust and pain facial expressions entail different trustworthiness processing, our Study 1 results seem to indicate a difference, at least for real facial expressions. We tentatively speculate that an evolutionary account of disgust and pain communication might offer one possible mechanism underpinning the lower trustworthiness reported by onlookers (Steinkopf, 2016). Individuals observing disgust, may interpret it as a cue of a threat or potential harm, thus inducing them to withdraw from interaction. Less straightforwardly, because pain serves simultaneously as a signal of potential threat to others and as a request for help and care by others, it poses some significant interpretative challenges to the observer's mind. Indeed, especially when contextual cues are poor or lacking (e.g., absence of a visible wound), the sufferer might be exaggerating (even faking) pain (Finlay & Syal, 2014), thus potentially posing a threat themselves or exploiting the observer's assistance (Steinkopf, 2015; Williams, 2002). It follows that identifying pain expressions as untrustworthy may depend more than disgust expressions on contextual information. This is a question yet to be addressed by empirical research.

Although unclear whether computer-generated faces are judged as less trustworthy than real faces (Balas & Pacella, 2017; but see Nightingale & Farid, 2022 for recent antithetical outcomes), previous research already highlighted the neural and perceptual difference between real and artificial faces (Balas et al., 2018), or even real but posed facial expressions (Jia et al., 2021), thus suggesting the presence of an intrinsic role of the "reality" and "naturalness" of the stimulus. However, when assessing posed and genuine expressions of pain, participants do not seem to make a significant distinction (Mende-Siedlecki et al., 2020 for a critical assessment) or even attribute more intense pain to posed expressions (Fernandes-Magalhaes et al., 2022).

4.1. Limitations and future directions

Based on current results, we would rather refrain from further speculating on the potential difference between disgust and pain, for two main reasons. First, Study 2 sample is small, and replication with a much larger sample would be necessary. Second, demand characteristics and social desirability phenomena may explain the lower trust assigned to negative expressions in both studies, particularly in Study 1, where respondents did not receive the instruction to respond as fast as possible. Our respondents might have answered what researchers were expecting them to answer, thus contaminating the interpretation of our findings (de Quidt et al., 2018; Mummolo and Peterson, 2019). This explanation may be more likely (or sufficient) for computer-generated expressions due to the near-ceiling categorization performance observed compared with real expressions (i.e., perfect imputation of trustworthiness to happy and neutral expressions and untrustworthiness to disgust and pain). Nevertheless, to improve the interpretability of results, future studies might favor a web-survey task-based experimental approach (instead of the correlational design we used in Study 1). Finally, we must also acknowledge that 1) we compared pain expression only with 1 positive and 1 negative type of expression, 2) the stimuli had little ecological validity because they were bidimensional static stimuli devoid of contextual information, 3) we had a prevalence of female participants that displayed a significantly biased judgment and performance depending on the sex of the actors displaying the real expressions (cf. interaction effects with sex in supplementary Appendixes B and D, Tables S2, S6). All in all, diversifying the sample (moving beyond the predominance of female and student participants), and increasing the sample size and the number of trials (perhaps with an even more ecological stimulus material), would allow addressing the fine-grained question of whether there is a difference in trustworthiness judgments/categorization performance during observation of pain vs disgust expressions.

To replicate the current findings, researchers will also have to consider other methodological nuances we overlooked. For example, we used a single neutral computer-generated face to reduce the number of confounding variables, such as age and race, and thus avoiding complicating the design. However, this choice led us to having a limited number of stimuli for the computer-generated set compared with the real faces set because the latter would vary per gender and actor (i.e., $n = 24$; neutral expressions: $n = 6$). By the same token, the habituation caused by single neutral expressions for the computer-generated faces may have been paralled by similar habituation for the real expressions, for which the different degrees of expression were not possible. Besides, the fact that computer-generated pain and disgust facial expressions were found different only when 50 % degree of expression was included in the analysis suggests that, even though participants successfully distinguished between emotions, the ambiguity/dynamics of expression may still affect their trustworthiness judgments for negative expressions (Appendix D, Table S9). Future studies investigating the relationship between perceived facial trustworthiness (Todorov et al., 2015 for a review) and pain display should consider the limitations mentioned above. Nevertheless, we believe our findings may spark a new stream of research to assess the perceptual and cognitive determinants of trustworthiness associated with facial expressions of pain in the lab and clinical settings. For example, future research may integrate advancements in artificial intelligence (AI), which hint at increased realisticness of AI-synthesized faces and increased trustworthiness (compared with

real faces, see Nightingale & Farid, 2022).

5. Conclusions

Our findings reveal for the first time that both pain and disgust expressions were perceived as less trustworthy compared with positive expressions. This outcome is consistent with overgeneralization for positive vs negative facial expressions. In addition, computer-generated facial expressions of pain are perceived as untrustworthy as disgust expressions (compared with positive facial expressions). Perceptual and cognitive biases during the trustworthiness judgment of a patient's facial expression may significantly contribute to chronic pain stigmatization. Thus, the current study provides evidence for an early perceptual mechanism (face perception) that may play a role in distrusting the patient's pain reports by observers and clinicians and warrant more research in both experimental and clinical settings.

Declaration of competing interest

The authors declare that there is no conflict of interest.

Data availability

As per information in the methods, data and relevant material are available on OSF repository.

Acknowledgements

M.V.D.B thanks the Research Foundation Flanders (FWO) for supporting this study (11K2721N, aspirant fundamental research fellowship). E.C. was supported by a senior postdoctoral fellowship awarded by the Research Foundation Flanders (12U0322N). E.V. thanks Emilia Ilieva for helping with initial set-up of Study 2.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2023.103893.

References

Albohn, D. N., Brandenburg, J. C., & Adams, R. B. (2019). Perceiving emotion in the "Neutral" face: A powerful mechanism of person perception. In U. Hess, & S. Hareli (Eds.), *The social nature of emotion expression*. Cham: Springer. https://doi.org/10.1007/978-3-030-32968-6_3.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Balas, B., & Pacella, J. (2017). Trustworthiness perception is disrupted in artificial faces. *Computers in Human Behavior, 77*, 240–248. https://doi.org/10.1016/j.chb.2017.08.045

Balas, B., Tupa, L., & Pacella, J. (2018). Measuring social variables in real and artificial faces. *Computers in Human Behavior, 88*, 236–243. https://doi.org/10.1016/j.chb.2018.07.013

Banaji, M. R., Fiske, S. T., & Massey, D. S. (2021). Systemic racism: Individuals and interactions, institutions and society. *Cognitive Research: Principles and Implications, 6*(1), 82. https://doi.org/10.1186/s41235-021-00349-3

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). Effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software, 5*(56), 2815. https://doi.org/10.21105/joss.02815

Blomqvist, K., & Edberg, A.-K. (2002). Living with persistent pain: Experiences of older people receiving home care. *Journal of Advanced Nursing, 40*(3), 297–306. https://doi.org/10.1046/j.1365-2648.2002.02371.x

Clarke, K. A., & Iphofen, R. (2005). Believing the patient with chronic pain: A review of the literature. *British Journal of Nursing (Mark Allen Publishing), 14*(9), 490–493. https://doi.org/10.12968/bjon.2005.14.9.18073

Couch, L. L., & Jones, W. H. (1997). Measuring levels of trust. *Journal of Research in Personality, 31*(3), 319–336. https://doi.org/10.1006/jrpe.1997.2186

de Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review, 108*(11), 3266–3302. https://doi.org/10.1257/aer.20171330

De Ruddere, L., & Craig, K. D. (2016). Understanding stigma and chronic pain: A-state-of-the-art review. *Pain, 157*(8), 1607–1610. https://doi.org/10.1097/j.pain.0000000000000512

De Ruddere, L., Goubert, L., Stevens, M. A. L., Deveugele, M., Craig, K. D., & Crombez, G. (2014). Health care professionals' reactions to patient pain: Impact of knowledge about medical evidence and psychosocial influences. *The Journal of Pain, 15*(3), 262–270. https://doi.org/10.1016/j.jpain.2013.11.002

De Ruddere, L., Goubert, L., Vervoort, T., Prkachin, K. M., & Crombez, G. (2012). We discount the pain of others when pain has no medical explanation. *The Journal of Pain, 13*(12), 1198–1205. https://doi.org/10.1016/j.jpain.2012.09.002

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science, 17*(5), 383–386. https://doi.org/10.1111/j.1467-9280.2006.01716.x

Ekman, P., & Friesen, W. V. (1978). *Manual for the facial action code*. Consulting Psychologist Press.

Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation. *Perception, 39*(7), 931–941. https://doi.org/10.1068/p6633

Fernandes-Magalhaes, R., Carpio, A., Ferrera, D., Van Ryckeghem, D., Peláez, I., Barjola, P., De Lahoz, M. E., Martín-Buro, M. C., Hinojosa, J. A., Van Damme, S., Carretié, L., & Mercado, F. (2022). Pain E-motion faces database (PEMF): Pain-related micro-clips for emotion research. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01992-4

Finlay, B. L., & Syal, S. (2014). The pain of altruism. *Trends in Cognitive Sciences, 18*(12), 615–617. https://doi.org/10.1016/j.tics.2014.08.002

Franklin, R. G., & Zebrowitz, L. A. (2013). Older adults' trait impressions of faces are sensitive to subtle resemblance to emotions. *Journal of Nonverbal Behavior, 37*(3), 139–151. https://doi.org/10.1007/s10919-013-0150-4

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*(1), 226–241. https://doi.org/10.3758/BRM.42.1.226

Freeman, J. B., Nakayama, K., & Ambady, N. (2013). Finger in flight reveals parallel categorization across multiple social dimensions. *Social Cognition, 31*(6), 792–805. https://doi.org/10.1521/soco.2013.31.6.792

Freitag, M., & Traunmüller, R. (2009). Spheres of trust: An empirical analysis of the foundations of particularised and generalised trust. *European Journal of Political Research, 48*(6), 782–803. https://doi.org/10.1111/j.1475-6765.2009.00849.x

Hale, J., Payne, M. E., Taylor, K. M., Paoletti, D., & De C Hamilton, A. F. (2018). The virtual maze: A behavioural tool for measuring trust. *Quarterly Journal of Experimental Psychology, 71*(4), 989–1008. https://doi.org/10.1080/17470218.2017.1307865

Hehman, E., Leitner, J. B., Deegan, M. P., & Gaertner, S. L. (2015). Picking teams: When dominant facial structure is preferred. *Journal of Experimental Social Psychology, 59*, 51–59. https://doi.org/10.1016/j.jesp.2015.03.007

Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proceedings of the National Academy of Sciences of the United States of America, 113*(16), 4296–4301. https://doi.org/10.1073/pnas.1516047113

Jia, S., Wang, S., Hu, C., Webster, P. J., & Li, X. (2021). Detection of genuine and posed facial expressions of emotion: Databases and methods. *Frontiers in Psychology, 11*, 3818. https://doi.org/10.3389/fpsyg.2020.580287

Johnson-George, C., & Swap, W. C. (1982). Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. *Journal of Personality and Social Psychology, 43*, 1306–1317. https://doi.org/10.1037/0022-3514.43.6.1306

Kätsyri, J., de Gelder, B., & de Borst, A. W. (2020). Amygdala responds to direct gaze in real but not in computer-generated faces. *NeuroImage, 204*, Article 116216. https://doi.org/10.1016/j.neuroimage.2019.116216

Katz, J., Rosenbloom, B. N., & Fashler, S. (2015). Chronic pain, psychopathology, and DSM-5 somatic symptom disorder. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie, 60*(4), 160–167. https://doi.org/10.1177/070674371506000402

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods, 49*(5), 1652–1667. https://doi.org/10.3758/s13428-017-0900-z

Kugler, T., Ye, B., Motro, D., & Noussair, C. N. (2020). On trust and disgust: Evidence from face reading and virtual reality. *Social Psychological and Personality Science, 11*(3), 317–325. https://doi.org/10.1177/1948550619856302

Kunz, M., Peter, J., Huster, S., & Lautenbacher, S. (2013). Pain and disgust: The facial signaling of two aversive bodily experiences. *PLoS One, 8*(12), Article e83277. https://doi.org/10.1371/journal.pone.0083277

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lee, E., Kang, J. I., Park, I. H., Kim, J., & An, S. K. (2008). Is a neutral face really evaluated as being emotionally neutral? *Psychiatry Research, 157*(1–3), 77–85. https://doi.org/10.1016/j.psychres.2007.02.005

Maldonado, M., Dunbar, E., & Chemla, E. (2019). Mouse tracking as a window into decision making. *Behavior Research Methods, 51*(3), 1085–1101. https://doi.org/10.3758/s13428-018-01194-x

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mende-Siedlecki, P., Qu-Lee, J., Lin, J., Drain, A., & Goharzad, A. (2020). The Delaware pain database: A set of painful expressions and corresponding norming data. *Pain Reports, 5*(6), Article e853. https://doi.org/10.1097/PR9.0000000000000853

Montali, L., Monica, C., Riva, P., & Cipriani, R. (2011). Conflicting representations of pain: A qualitative analysis of health care professionals' discourse. *Pain Medicine (Malden, Mass.), 12*(11), 1585–1593. https://doi.org/10.1111/j.1526-4637.2011.01252.x

Mummolo, J., & Peterson, E. (2019). Demand effects in survey experiments: An empirical assessment. *American Political Science Review, 113*(2), 517–529. https://doi.org/10.1017/S0003055418000837

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences, 119*(8), Article e2120481119. https://doi.org/10.1073/pnas.2120481119

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion (WashingtonD.C.), 9*(1), 128–133. https://doi.org/10.1037/a0014520

Patrick, C. J., Craig, K. D., & Prkachin, K. M. (1986). Observer judgments of acute pain: Facial action determinants. *Journal of Personality and Social Psychology, 50*(6), 1291–1298. https://doi.org/10.1037//0022-3514.50.6.1291

Prkachin, K. M. (1992). The consistency of facial expressions of pain: A comparison across modalities. *Pain, 51*(3), 297–306. https://doi.org/10.1016/0304-3959(92)90213-U

R Core Team. (2021). R: A Language and envionment for statistical computing. https://cran.r-project.org/.

Riva, P., Rusconi, P., Montali, L., & Cherubini, P. (2011). The influence of anchoring on pain judgment. *Journal of Pain and Symptom Management, 42*(2), 265–277. https://doi.org/10.1016/j.jpainsymman.2010.10.264

Riva, P., Sacchi, S., Montali, L., & Frigerio, A. (2011). Gender effects in pain detection: Speed and accuracy in decoding female and male pain expressions. *European Journal of Pain (London, England), 15*(9), 985.e1–985.e11. https://doi.org/10.1016/j.ejpain.2011.02.006

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26*(5), 443–452. https://doi.org/10.1037/h0031464

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology, 104*(3), 409–426. https://doi.org/10.1037/a0031050

Rusconi, P., Riva, P., Cherubini, P., & Montali, L. (2010). Taking into account the observers' uncertainty: A graduated approach to the credibility of the patient's pain evaluation. *Journal of Behavioral Medicine, 33*(1), 60–71. https://doi.org/10.1007/s10865-009-9232-5

Simon, D., Craig, K. D., Gosselin, F., Belin, P., & Rainville, P. (2008). Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *Pain, 135*(1), 55–64. https://doi.org/10.1016/j.pain.2007.05.008

Sims, O. T., Gupta, J., Missmer, S. A., & Aninye, I. O. (2021). Stigma and endometriosis: A brief overview and recommendations to improve psychosocial well-being and diagnostic delay. *International Journal of Environmental Research and Public Health, 18*(15), 8210. https://doi.org/10.3390/ijerph18158210

Steinkopf, L. (2015). The signaling theory of symptoms: An evolutionary explanation of the placebo effect. *Evolutionary Psychology, 13*(3). https://doi.org/10.1177/1474704915600559, 1474704915600559.

Steinkopf, L. (2016). An evolutionary perspective on pain communication. *Evolutionary Psychology, 14*(2). https://doi.org/10.1177/1474704916653964, 1474704916653964.

Sutherland, C. A. M., Rowley, L. E., Amoaku, U. T., Daguzan, E., Kidd-Rossiter, K. A., Maceviciute, U., & Young, A. W. (2015). Personality judgments from everyday images of faces. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.01616

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience, 3*(2), 119–127. https://doi.org/10.1093/scan/nsn009

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology, 25*(3), 395–410. https://doi.org/10.1080/02643290802044996

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science (New York, N.Y.), 308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Tuck, N. L., Johnson, M. H., & Bean, D. J. (2019). You'd better believe it: The conceptual and practical challenges of assessing malingering in patients with chronic pain. *The Journal of Pain, 20*(2), 133–145. https://doi.org/10.1016/j.jpain.2018.07.002

Ueda, Y., Nagoya, K., Yoshikawa, S., & Nomura, M. (2017). Forming facial expressions influences assessment of others' dominance but not trustworthiness. *Frontiers in Psychology, 8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.02097.

Wakefield, E. O., Puhl, R. M., Litt, M. D., & Zempsky, W. T. (2021). 'If it ever really hurts, I try not to let them know:' The use of concealment as a coping strategy among adolescents with chronic pain. *Frontiers in Psychology, 12*, Article 666275. https://doi.org/10.3389/fpsyg.2021.666275

Wells, N., Pasero, C., & McCaffery, M. (2008). Improving the quality of care through pain assessment and management. In R. G. Hughes (Ed.), *Patient Safety and Quality: An Evidence-based Handbook for Nurses*. US: Agency for Healthcare Research and Quality. http://www.ncbi.nlm.nih.gov/books/NBK2658/.

Williams, A. C.d. C. (2002). Facial expression of pain: An evolutionary account. *Behavioral and Brain Sciences, 25*(4), 439–455. https://doi.org/10.1017/S0140525X02000080

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*(3), 1497. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Zgonnikov, A., Aleni, A., Piiroinen, P. T., O'Hora, D., & di Bernardo, M. (2017). Decision landscapes: Visualizing mouse-tracking data. *Royal Society Open Science, 4*(11), Article 170482. https://doi.org/10.1098/rsos.170482