# Sex differences in emotion recognition: evidence for a small overall female superiority on facial disgust

Hannah L. Connolly[1]*, Carmen E. Lefevre[2], Andrew W. Young[3], and Gary J. Lewis[1]

1. Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK
2. Centre for Behaviour Change, Department of Clinical, Educational and Health Psychology, University College London, 1-19 Torrington Place, University College London, London, WC1E 7HB, UK
3. Department of Psychology, University of York, Heslington, York YO10 5DD, UK

* Corresponding author email: hannah.connolly.2016@live.rhul.ac.uk

Revision submission date: 21/02/18

## Abstract

Although it is widely believed that females outperform males in the ability to recognise other people's emotions, this conclusion is not well supported by the extant literature. The current study sought to provide a strong test of the female superiority hypothesis by investigating sex differences in emotion recognition for five basic emotions using stimuli well-calibrated for individual differences assessment, across two expressive domains (face and body), and in a large sample (N=1022: Study 1). We also assessed the stability and generalisability of our findings with two independent replication samples (N=303: Study 2, N=634: Study 3). In Study 1, we observed that females were superior to males in recognising facial disgust and sadness. In contrast, males were superior to females in recognising bodily happiness. The female superiority for recognition of facial disgust was replicated in Studies 2 and 3, and this observation also extended to an independent stimulus set in Study 2. No other sex differences were stable across studies. These findings provide evidence for the presence of sex differences in emotion recognition ability, but show that these differences are modest in magnitude and appear to be limited to facial disgust. We discuss whether this sex difference may reflect human evolutionary imperatives concerning reproductive fitness and child-care.


**Keywords**: emotion recognition; face; body; sex differences; disgust

## Introduction

The ability to accurately recognise other people's emotions is a core socio-cognitive skill (Bruce & Young, 1986, 2012; Young, 2016). Although much work in this domain has emphasised that emotion recognition ability is to some extent akin to an innate human faculty (Darwin, 1872), such that certain emotions are universally recognised, regardless of culture (Ekman & Friesen, 1971), it is also clear that not all people can recognise emotional expressions equally well (Lewis, Lefevre, & Young, 2016). With regard to these individual differences, one of the most widely discussed factors that may influence emotion recognition ability is biological sex (Kret & De Gelder, 2012), where meta-analytic work has claimed that on average women outperform men (Hall, 1978; McClure, 2000; Thompson & Voyer, 2014).

However, despite the widely-held assertion of female 'superiority' in emotion recognition ability, a closer examination of the literature reveals a more mixed picture, as detailed below. Furthermore, the literature is almost entirely reliant on studies of facial expression recognition. In other expressive domains, such as recognition of emotion from body postures, little is known regarding sex differences. This warrants further work, as emotion recognition ability has important real-world implications: accurate recognition of emotions is associated with better social functioning (Brackett et al., 2006), whereas recognition difficulties are linked to interpersonal problems and to the aetiology and maintenance of several psychiatric disorders, such as depression (Surguladze et al., 2004). Given the different prevalence rates of psychopathology across the sexes, sex differences in basic socio-cognitive processes (such as emotion recognition ability) which may contribute to such outcomes (e.g. in depression: Harmer, Goodwin & Cowen, 2009) are thus important phenomena to delineate and study. Moreover, different evolutionary and socialisation niches occupied by males and females make the study of sex differences important for basic scientific enquiry.

**Sex Differences in Facial Emotion Recognition – A Brief Overview**

Much of the earlier work in the field has investigated emotion recognition as a general ability based on a global score rather than distinguishing between individual emotions (e.g. Buck, Miller, & Caul, 1974; Kirouac & Dore, 1985). More recent work instead tests recognition of affect more specifically through the use of a wider range of individually-scored basic emotions. In line with this approach, here we explicitly focus on recent research examining the five basic emotions of anger, disgust, fear, happiness, and sadness.

In a study addressing recognition ability for negatively-valenced basic emotions (anger, disgust, fear, and sadness), Rotter and Rotter (1988) ($N_{Study1}$=727: 214 males; $N_{Study2}$=399: 162 males) noted a female advantage across all expressions. Hall and Matsumoto (2004), across two studies with different stimuli exposure times (Study 1: 10s, Study 2: 200ms) and using five basic emotions, found that females were significantly more accurate for disgust, happiness, and sadness (Study 1: N=96: 69 males), and for anger, disgust, fear, and sadness (Study 2: N=363: 126 males). Lee and colleagues (2013), using a large sample of adolescents (N=1954: 956 males), found a female advantage for discriminating morphed facial expressions on the continua of Happiness-Fear and Happiness-Sadness (but not the Anger-Sadness or Anger-Fear continua). More recently, Duesenberg and colleagues (2016) tested 80 participants (40 males) on two emotions, anger and sadness, each presented at two intensities of expression, 40% and 80%. These authors reported no sex differences for the sadness stimuli, but found a significant female advantage for accurately identifying anger at both intensities.

However, in contrast to these studies that were broadly supportive of a female advantage in emotion recognition, a number of studies have reported either no sex differences or even a male advantage. Rahman, Wilson and Abrahams (2004) assessed 240 participants

(120 males) on happiness and sadness, but reported no accuracy advantage for either emotion. Grimshaw, Bulman-Fleming, and Ngo (2004) (N=73: 36 males) presented stimuli for 50ms portraying three basic emotions of anger, happiness, and sadness, but found no sex difference in accuracy or reaction times. Unlike many other studies in this area which used a forced choice paradigm, Williams and Mattingley (2006) tested 156 (78 males) participants on rapid detection of angry or fearful faces amongst neutral distractors. Males were found to be significantly faster at identifying angry male faces, but there was no difference in speed between sexes to detect fearful stimuli. Again using a visual search paradigm, Sawada and colleagues (2014) (N=90: 46 males) measured participant ability in detecting either an angry or happy face amongst other neutral faces, and found no significant effect of sex on reaction time. Testing a large sample of undergraduates (N=993: 211 males) on facial expressions in both frontal and profile views, Matsumoto and Hwang (2011) did not observe significant sex differences in ability to categorise the five basic emotions. Most recently, Lyusin and Ovsyannikova (2016) tested a large sample of participants (N=684: 221 males) on recognition accuracy and sensitivity on 15 different emotions (including the five basic emotions) using naturalistic video recordings, but reported no significant difference between the sexes on these measures.

**Sex Differences in Bodily Emotion Recognition – A Brief Overview**

Bodily emotion recognition is the ability to distinguish a portrayed emotion from the stimulus' body form alone. It is a fundamental component of accurate emotion perception, with research showing that body posture can be critical in resolving ambiguous facial expressions (Aviezer, Trope, & Todorov, 2012). The ability to accurately perceive emotion from both face and body may therefore represent an optimal strategy for emotion recognition, in that all the available cues are integrated to form the most informed interpretation (Young

& Bruce, 2011; Young, 2018). This strategy seems especially pertinent in situations in which one channel is occluded or ambiguous, or when moving face and body signals are expressed very rapidly.

Body emotion recognition ability can be assessed through both static and dynamic stimuli. In the case of static stimuli, a photograph of a person with their face occluded and expressing a given emotion is presented to the participant. Dynamic stimuli often consist of point-light displays showing a set of locations on a human body making a natural movement (see Johansson, 1973, and Figure 1, or the video in the Supplementary Materials). Whereas a large body of research has addressed sex differences in facial emotion recognition, only a handful of studies to date have addressed sex differences in body emotion recognition. An early study by Sogon and Izard (1987) (N=94: 47 males) involved short video clips of scenes portraying five emotions (surprise, contempt, affection, anticipation, or acceptance), and found females to be significantly better at identifying disgust and fear. Additionally their ability to recognise sadness also approached significance (Sogon & Izard, 1987). A more recent study (N=37: 15 males) involving happy, sad, and angry body stimuli found females to be faster in recognising these emotions from point-light displays, but found no significant sex difference in overall accuracy rates (Alaerts et al., 2011). Two further studies used happy and angry point-light displays, and observed a significant superiority for males in identifying happiness, as well as a tendency, albeit non-significant, for females to perform better at identifying anger (Sokolov et al., 2011; Krüger et al., 2013). However, it should be noted that all of these more recent studies used small samples (N<100) and assessed only two (Sokolov, 2011; Krüger, 2013) or three (Alaerts, 2011) of the basic emotions.

## The Current Study

To summarise the literature, a wide variety of claims have been made regarding sex differences in face and body emotion recognition ability. Despite meta-analytic work reporting an overall female advantage (Hall, 1978; McClure, 2000; Thompson & Voyer, 2014), a number of studies – some with relatively large samples (e.g. Lyusin & Ovsyannikova, 2016; Matsumoto & Hwang, 2011) – have not reported this pattern. Various studies reported sex differences only for specific emotion(s), and limited consensus can be reached given the nature of the extant literature.

Several factors may explain these mixed findings. Firstly, sample sizes have tended to be modest (N<200) and as such may be underpowered to detect what are unlikely to be large effect sizes. For example, Thompson and Voyer (2014) reported a Cohen's *d* of .19 in favour of females on emotion recognition tasks and a suitably powered test for an effect of this magnitude would require several hundred participants. Secondly, methodological factors may play a role, as widely used stimulus sets have often been developed with a view to creating easily recognised expressions. In consequence, a number of published studies show clear ceiling effects for their stimuli, which will diminish the possibility of detecting group differences (e.g. Hampson, van Anders, & Mullin, 2006; Hoffmann, Kessler, Eppel, Rukavina, & Traue, 2010). Thirdly, while a recent meta-analysis reported evidence for a generalised sex difference in emotion recognition ability (Thompson & Voyer, 2014), evidence of publication bias (as assessed by the Test of Excess Significance method proposed by Ioannidis & Trikalinos, 2007) was apparent. Finally, in the specific case of body emotion recognition, too few studies have been performed to gain traction on possible sex differences.

The present work therefore examined whether sex difference are present in emotion recognition ability, and if so, whether this sex difference is restricted to a specific emotion (across five basic emotions) or expressive domain (i.e., face or body stimuli). With the aforementioned issues in mind, our study made a contribution to the literature in four important ways. Firstly, we used a large sample of adults (N=1022: Study 1) in order to provide adequate statistical power to detect even modest group differences. Second, we used carefully developed stimulus sets – piloted prior to the current investigation – that do not show floor or ceiling effects and thus enhance the power to detect group differences (see Methods and Lewis et al., 2016 for further details). Thirdly, we used both face and body stimuli in order to examine if sex differences are restricted to one expressive domain or instead reflect more general processes. Finally, and with recent discussions of reproducibility of some findings in psychology (Open Science Collaboration, 2015) very much in mind, we used two further independent participant datasets (N=303: Study 2; N=634: Study 3) and additional different sets of face and body stimuli (Study 2) to test the robustness of any effects observed.

## Study 1

## Methods

### Participants

1063 participants were recruited through Amazon's MTurk service as part of a previous study that did not analyse sex differences in the obtained data (Lewis et al., 2016: Studies 1 and 2). Ethical approval for this study was granted by the Department of Psychology Ethics Committee at the University of York. As expected with an online presentation, a number of participants experienced technical failures (e.g. stimuli not displaying properly). Consequently, and in line with our previous data exclusion strategy we only included participants in our analyses who completed at least 90% (≥18 of 20) of trial blocks for each emotion (anger, disgust, fear, happiness, sadness) and expressive domain (face and body). We also excluded participants for whom responses indicated low attention (e.g. using the same response key repeatedly) and those who did not disclose their sex (N=7).

This led to the omission of 41 participants and a final sample size of 1022 (322 males). The gender ratio was near-identical across White and non-White participants. In order to detect the effect size reported in the most recent meta-analysis (Thompson & Voyer, 2014), our sample size provided power of .80 for a Cohen's *d* of .19 with a two-tailed *t* test and an alpha level of .05 (Faul, Erdfelder, Lang, & Buchner, 2007). Note, this power analysis reflects an allocation ratio of 2.2: i.e. there were just over twice as many females than males in this sample. The mean age of the sample was 36.2 years (SD = 12.0). A range of ethnicities were reported in the final sample: White (n=775), Hispanic (n=47), Asian (n=53), Black (n=30), Native American (n=11), Other (n=74), and Undisclosed (n=32). These demographics are typical for MTurk samples (Paolacci, Chandler, & Ipeirotis, 2010).

**Stimuli**

Examples of each of the stimulus sets are shown in Figure 1. Examples of the two dynamic stimulus sets are presented in a video file in the Supplementary Materials.

***Face stimuli (static)*****:** To capture individual differences in facial expression recognition abilities we used static image stimuli taken from the Facial Expressions of Emotion: Stimuli and Tests (FEEST) set (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002). In brief, a total of 10 identities each posing five basic emotions (anger, disgust, fear, happiness, and sadness) were selected from the Ekman and Friesen series of Pictures of Facial Affect (Ekman & Friesen, 1976). In order to avoid floor/ceiling effects, we piloted examples of each emotional expression morphed relative to the neutral expression of the same identity using Psychomorph (Tiddeman, Burt, & Perrett, 2001). This procedure is known to lead to changes in the perceived intensity (and hence recognisability) of emotion (Calder, Young, Rowland, & Perrett, 1997). Here it was used to create five intensities (25%, 50%, 75%, 100%, and 125%) of each prototype (100%) expression (total n=250 images). In a pilot experiment comprising undergraduate and postgraduate students at the University of York (n=12 participants: 4 males), we tested recognition accuracy for each of these stimuli in a five-alternative forced choice paradigm with a 1000ms exposure time. This step is of considerable importance as the limited scope of individual differences research on emotion recognition ability has meant that suitable stimuli (i.e., free of ceiling effects and with adequate variance for individual differences research) have usually been unavailable. We then selected sets of 10 stimuli for each emotion (i.e., total N=50) that showed adequate means and variances based on these pilot data (i.e., that were not showing clear floor or ceiling effects). In the cases where a surplus of stimuli was available, we chose 10 that varied in gender and age as much as possible. This selection approach was also applied to the remaining stimulus sets.

***Body Stimuli (dynamic)*:** To capture emotion recognition ability from body expressions we used dynamic point-light walker stimuli previously described by Atkinson, Dittrich, Gemmell, and Young (2004). In short, 10 actors were recorded performing each of five emotions at three levels of intensity (typical, exaggerated, very exaggerated). Actors wore suits with 13 reflective patches on key joints of their body. Subsequent rendering removed all information other than the patches from each video, resulting in a short clip of 13 light points whose combined movement simulated the natural, dynamic expression of a human emotion. Video clips lasted between 4.2 and 8 seconds. As with the face stimuli, we chose 10 stimuli for each emotion (i.e., total N=50) that showed adequate means and variances following a pilot experiment (n=12 participants: 6 males).

**----- Insert Figure 1 here -----**

**Procedure**

Stimuli were blocked according to expressive domain. Face and body blocks were each presented twice to the participants in a fixed order (face-body-face-body). In a five-alternative forced choice paradigm, participants had to select the emotion they thought was displayed by each stimulus using radio buttons on screen. Each face stimulus was presented for 1000ms. This limited exposure time was chosen in order to increase difficulty of the recognition task, and to ensure suitability of the data for individual differences research (i.e., by increasing error rates and variation in recognition accuracy). Body stimuli were presented for the duration of each video clip. Participants could provide their response at any point following the onset of the stimulus presentation. The within-block presentation order was fully randomised. Participants were given the opportunity to rest following completion of each block. The mean recognition performance across blocks for each emotion was used in our analyses.

**Analysis**

In order to assess possible sex differences in emotion recognition from face and body, we used the data from Study 1 to conduct a three-way repeated measures Analysis of Variance (ANOVA) (2×2×5) with sex (male, female) as a between-subjects factor, and expressive domain (face, body) and emotion (anger, disgust, fear, happiness, sadness) as repeated within-subjects measures. For significant interactions, we conducted post-hoc tests to further analyse the results. As noted earlier, the prior literature did not provide strong bases for hypothesis testing and so these analyses were exploratory in nature.

## Results

The data were submitted to a three-way ANOVA exploring the effects of sex, domain and emotion. There was a significant main effect of domain ($F(1, 1020) = 46.37$, $p<.001$, partial $\eta^2 = .04$ [CI95%: .02-.07]) with emotional expressions from faces being more accurately recognised (M=.61, SD=.17) than expressions from bodies (M=.59, SD=.18). For the factor of emotion, Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2$ (9) = 81.91, $p < .05$). Therefore, we corrected degrees of freedom using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .96$). The main effect of emotion was significant ($F(3.84, 3916.41) = 600.08$, $p<.001$, partial $\eta^2 = .37$ [CI95%: .35-.39]) indicating differential performance across the five emotions. Recognition accuracy was greatest for happiness (M=.69, SD=.16), followed by fear (M=.66, SD=.17), anger (M=.61, SD=.17), sadness (M=.59, SD=.18), and disgust (M=.45, SD=.19). The main effect of sex was non-significant ($F(1,1020) = 2.96$, $p=.086$, partial $\eta^2 = .003$ [CI95%: .00-.01]).

There was a significant interaction between expressive domain and sex ($F(1, 1020) = 33.32$, $p<.001$, partial $\eta^2 = .03$ [CI95%: .01-.06]). Further significant interactions were also found between domain and emotion ($F(3.78, 3859.82) = 1108.06$, $p<.001$, partial $\eta^2 = .52$ [CI95%: .50-.54]), and between sex and emotion ($F(3.84, 3916.41) = 6.28$, $p<.001$, partial $\eta^2 = .006$ [CI95%: .002-.011]). These interactions were also corrected using Greenhouse-Geisser estimates ($\varepsilon = .95$). Of importance, these main effects and two-way interactions were qualified by a significant three-way interaction (also Greenhouse-Geisser corrected) between expressive domain, emotion, and sex ($F(3.78, 3859.82) = 3.69$, $p=.006$, partial $\eta^2 = .004$ [CI95%: .0003-.007]), which we now explore in greater depth.

We examined the effects of emotion and sex for each domain separately, by running mixed ANOVAs with emotion as a within-subjects factor and sex as a between-subjects factor for face and body stimuli separately. In both domains, the emotion variable violated the assumption of sphericity, therefore Greenhouse-Geisser corrections were used ($\varepsilon_{Faces} = .95$,

$\varepsilon_{Bodies}$ = .91). For facial stimuli, there were significant main effects of emotion (F(3.82, 3891.79) = 762.72, p<.001, partial $\eta^2$ = .43 [CI95%: .41-.45]) and sex (F(1, 1020) = 25.03, p<.001, partial $\eta^2$ = .02 [CI95%: .009-.045]). These effects were qualified by a significant interaction between emotion and sex (F(3.82, 3891.79) = 2.99, p=.020, partial $\eta^2$ = .003 [CI95%: .0001-.006]). For body stimuli, the ANOVA revealed a significant main effect of emotion (F(3.64, 3717.38) = 902.14, p<.001, partial $\eta^2$ = .47 [CI95%: .45-.49]) and a significant interaction between emotion and sex (F(3.64, 3717.38) = 7.22, p<.001, partial $\eta^2$ = .007 [CI95%: .002-.013]) but no main effect of sex (F(1, 1020) = 1.91, p=.167, partial $\eta^2$ = .002 [CI95%: .00-.011]).

To probe the nature of each of the emotion and sex interactions we ran ten post hoc (Bonferroni-corrected: adjusted α = .005) *t* tests (i.e., one for each emotion across both expressive domains) comparing male and female performance (see Figure 2). These analyses revealed that females performed significantly better on recognition of facial expressions of disgust (*t*(1020) = 4.19, p<.001, Cohen’s *d* = .28 [CI95%: .15-.41]) (female M=.58, SD=.19; male M=.53, SD=.19), and sadness (*t*(1020) = 3.90, p<.001, Cohen’s *d* = .26 [CI95%: .13-.40]) (female M=.49, SD=.18; male M=.44, SD=.18), and that males performed significantly better for bodily expressions of happiness (*t*(1020) = 3.74, p<.001, Cohen’s *d* = .25 [CI95%: .12-.38]) (female M=.51, SD=.18; male M=.56, SD=.17). The remaining *t* tests were non-significant at our Bonferroni-corrected alpha level (all $t \leq 2.45$, all p > .02).

**----- Insert Figure 2 here -----**

## Discussion

Study 1 examined the presence of sex differences in recognising emotion across face and body stimuli in a large sample. The results indicated that whilst sex had no overall effect, significant interactions between sex, expressive domain, and emotion did emerge. Upon further inspection of the nature of these interactions, we observed a significant female advantage for recognising facial disgust and facial sadness, and a significant male advantage for recognising bodily happiness. No other significant sex differences were noted.

We next sought to confirm the key findings observed from Study 1 in a replication sample. In addition, we sought to assess whether these findings extended to reflect emotion processing more broadly. To this end we used data from a sample of participants who had completed an emotion recognition battery involving the same tests reported above and two novel stimulus sets: specifically, dynamic facial stimuli and static body stimuli.

## Study 2

## Methods

### Participants

In Study 2 we analysed a set of archival data from an independent MTurk sample, also drawn from the same previous study (as reported in Lewis et al., 2016: Study 3, N=384). Again, we only included participants in our analyses who completed at least 90% (≥17 of 19: note, this study did not assess static bodily disgust – see below for further details) of the trial blocks for each emotion and expressive domain, and showed no evidence of false responding, leading to the omission of 81 participants and a final sample size of 303 (137 males). The gender ratio was near-identical across White and non-White participants. In order to test the specific effects observed in Study 1, this sample provided power of .69, .72, and .78 (for a Cohen's *d* of .25, .26, & .28, respectively) for a one-tailed *t* test with an alpha level of .05. Mean age was 34.8 years (SD = 11.3). A range of ethnicities were reported in the sample: White (n=232), Hispanic (n=16), Asian (n=16), Black (n=10), Native American (n=1), Other (n=15), and Undisclosed (n=13).

**Stimuli**

We used the same stimuli as described above, together with the additional sets of face and body stimuli detailed next. See Figure 1 for picture examples of each of our four types of stimuli, and the video in the Supplementary Materials for examples of the dynamic stimuli.

***Face stimuli (dynamic)***: We used a sub-set of dynamic facial stimuli previously used for emotion recognition work (Lau et al., 2009). In brief, these stimuli were created by morphing one male and one female image from a neutral expression to one of the five basic emotions (anger, disgust, fear, happiness, and sadness) and then assembling the morphed images into a video clip. Stimuli dynamically changed from the neutral expression to one of four levels of intensity (25%, 50%, 75%, 100%), but for happiness, due to the ceiling effects often observed, intensity levels were lower (10%, 25%, 50%, and 75%). We only used stimuli where actors directly faced the camera, with either direct or averted gaze. This led to a total of 80 stimuli that we piloted as before, in order to avoid floor/ceiling effects (n=47 participants: 28 males, recruited through Amazon MTurk) before selecting sets of 10 stimuli for each emotion (i.e., total N=50) that showed adequate means and variances based on these pilot data. Each video clip was approximately 1.5 seconds in length.

***Body stimuli (static)***: To test emotion recognition from static bodies we employed the Bodily Expressive Action Stimuli Test (BEAST) stimuli set (de Gelder & Van den Stock, 2011). In brief, these static stimuli comprised black and white whole body photographs of actors with faces obscured depicting one of four basic emotions (anger, fear, happiness, and sadness). Disgust is not included in this stimulus set due to it being difficult to represent in the static body alone (de Gelder & Van den Stock, 2011). The original image set contains 254 images. We again undertook undergraduate piloting at the University of York (n=14 participants: 6 males) to identify 10 stimuli per emotion (i.e., total N=40) suitable for an individual differences task, for which we then validated means and variance in a second pilot study using MTurk participants (n=50). As with the static facial images, we presented each image for 1000ms.

**Procedure**

The procedure was similar to that outlined for Study 1 with the exception that participants completed the following blocks in fixed order: static bodies, static faces, dynamic bodies, and dynamic faces. As such, each stimulus set was only seen once (as opposed to twice in Study 1). We describe Study 2 as a replication sample in the sense that we used the same stimulus sets in our second sample as we did in Study 1 (alongside two additional stimulus sets to test for the generalisability of the effects), as well as only examining the significant results that emerged in Study 1.

**Analysis**

Here we attempted to replicate only the significant effects observed in Study 1 to assess their robustness, as well as to examine whether these effects generalised to the additional stimulus sets. We chose to constrain our analyses to only the significant effects from Study 1, as any effects emerging in Study 2 that were not observed in Study 1 were unlikely to be of substantive interest given the much larger sample size of Study 1.

## Results

We ran 6 *t* tests based on the three significant findings from the discovery sample (i.e., facial disgust, facial sadness and bodily happiness) and tested each of these in both static and dynamic forms. Therefore, our six *t* tests comprised static facial disgust, dynamic facial disgust, static facial sadness, dynamic facial sadness, static bodily happiness, and dynamic bodily happiness.

As in Sample 1, females performed significantly better on recognition of static facial expressions of disgust (*t*(301) = 3.54, p<.001, Cohen's *d* = .41 [CI95%: .18-.64]) (female M=.61, SD=.21; male M=.52, SD=.21). In addition, they also performed better than males on recognition of dynamic facial expressions of disgust (*t*(301) = 4.09, p<.001, Cohen's *d* = .47 [CI95%: .24-.70]) (female M=.79, SD=.18; male M=.70, SD=.18).

In contrast to Study 1, females scored significantly higher on recognition of happiness from static bodies (*t*(301) = 2.09, p=.038, Cohen's *d* = .24 [CI95%: .01-.47]) (female M=.48, SD=.23; male M=.42, SD=.22). However, we observed no differences between males and females for dynamic expressions of bodily happiness (*t*(301) = .49, p=.625, Cohen's *d* = .06 [CI95%: -.17, .28]) (female M=.54, SD=.18; male M=.53, SD=.20). Finally, in contrast to Study 1 we observed no difference between males and females for either static facial expressions of sadness (*t*(301) = 1.09, p=.275, Cohen's *d* = .13 [CI95%: -.10, .35]) (female M=.48, SD=.23; male M=.46, SD=.21) or for dynamic facial expressions of sadness (*t*(301) = .21, p=.835, Cohen's *d* = .02 [CI95%: -.20, .25]) (female M=.43, SD=.27; male M=.42, SD=.25). These results are also detailed in Figure 3.

**----- Insert Figure 3 here -----**

## Discussion

The results from Study 2 confirm the significant female advantage for facial disgust that was observed in Study 1. Of importance, this sex difference in disgust recognition was also observed in an additional stimulus set comprising dynamic facial stimuli. In addition, females scored significantly higher on recognition of static bodily happiness, although this should be interpreted in the context of a significant male advantage in Study 1. No other sex differences emerged.

Although these results confirm the finding of Study 1 with regards to sex differences in disgust recognition ability, our statistical power was lower than conventional levels (i.e. below 80%). This will have limited our ability to detect sex differences in facial sadness and bodily happiness, as well as raised the potential for a false positive in our disgust recognition observation. With this in mind we conducted a third study with a sample size adequately powered to reliably detect effects of the magnitude we observed in Study 1.

## Study 3

## Methods

### Participants

In order to assess the robustness of the effects observed in Studies 1 and 2, we collected a third independent sample of MTurk participants. We sought to recruit a sample that would provide power of .80 to detect a Cohen's *d* of .20 in a one-tailed test with alpha at .05. This level of power was chosen with the concern that the previously observed effect sizes – particularly of Sample 1 with its larger sample – might be overestimates of the population effect size. To this end we continued recruiting until we had usable data that satisfied this power requirement (note, because the required sample size for a given level of power is sensitive to the gender ratio, and because we of course could not precisely know this ratio ahead of time for our final sample, we assessed the power of our sample – with our data exclusion protocol in place – periodically throughout recruitment). As before, we only included participants in our analyses if they completed 90% (≥9 of 10) of the trial blocks for each emotion and expressive domain, and showed no evidence of false responding. In total 730 participants completed our survey. 96 participants were omitted following our data exclusion protocol. Our final sample size was thus N=634 (275 males) and this sample provided power of exactly .80 to detect a Cohen's *d* of .20 in a one-tailed test with alpha at

.05. The gender ratio was near-identical across White and non-White participants. The mean age was 36.8 years (SD=10.9), and ethnicity was reported as follows: White (n=490), Hispanic (n=34), Asian (n=38), Black (n=59), Native American (n=3), Other (n=8), and Undisclosed (n=2).

**Stimuli**

For this sample, we used the same stimuli as detailed in Study 1; specifically there were 10 static faces and 10 dynamic bodies for each of the five basic emotions (i.e., N=50 for each expressive domain).

**Procedure**

The procedure for Study 3 was the same as in Study 1, with the exception that each block was presented only once to the participants (as opposed to being presented twice). Participants completed the two blocks in the same fixed order: static faces, dynamic bodies. Within-block presentation order was fully randomised. The stimuli presentation and response procedure was the same as outlined in the previous two samples.

**Analysis**

Here we attempted to replicate only the significant effects observed in Study 1 to assess their robustness.

## Results

We ran 3 *t* tests based on the three significant findings that emerged in Study 1 (i.e., facial disgust, facial sadness, and bodily happiness). The tests confirmed a significant female advantage for static facial disgust ($t(630) = 2.21$, p=.027, Cohen's $d = .18$ [CI95%: .02-.33]) (female M=.60, SD=.20; male M=.56, SD=.21). There was no significant sex difference in recognition of facial sadness (*t*(630) = 1.22, p=.223, Cohen's *d* = .10 [CI95%: -.06, .26]) (female M=.46, SD=.20; male M=.44, SD=.20), or bodily happiness (*t*(630) = 1.19, p=.234, Cohen's *d* = .10 [CI95%: -.06, .25]) (female M=.48, SD=.18; male M=.50, SD=.18) (see Figure 4).

**----- Insert Figure 4 here -----**

## Discussion

The results of Study 3 confirm the female advantage for recognising facial disgust observed in both Studies 1 and 2. We did not see evidence for a sex difference in facial sadness and bodily happiness.

## General Discussion

The current study sought to determine the role of sex in emotion recognition ability across three large samples of adults. In Study 1 (our discovery sample), while no overall main effect of sex was observed, significant interactions across sex, expressive domain, and emotion were noted. Following decomposition of these interactions, we observed that females performed significantly better than males on recognition of facial disgust and facial sadness, and that males performed significantly better than females in recognising bodily happiness. Results from Study 2 (our first replication sample) confirmed the female advantage for facial disgust. Importantly, this significant difference was present both for static disgust stimuli (as used in Studies 1 and 2) and for dynamic facial disgust stimuli (used only in Study 2). Results from Study 3 (our second replication sample) also confirmed the female advantage for static facial disgust. This pattern of findings shows that across three independent samples, with different presentation formats (static or dynamic), females consistently outperformed males with regard to recognising facial disgust stimuli.

In contrast to this consistent female advantage for recognising facially expressed disgust, the findings in Study 1 of a female advantage for sad faces and of a male advantage for happy bodies were not replicated. In fact, in Study 2 we observed the opposite effect: females were significantly better at recognising happiness from static bodies (with no sex differences observed for dynamic happy bodies), and in Study 3, neither facial sadness nor bodily happiness showed significant sex differences. These results indicate that the findings in Study 1 probably reflect random sampling variability and so we do not discuss these failures to replicate any further.

In sum, then, these findings support the existence of a robust sex difference for facial disgust but not for any of the other basic emotions. It should be noted, however, that the effect sizes observed are not large, and thus the distributions of scores for male and female participants are largely overlapping. As such, the sex difference should not be overstated; the two sexes appear to be more similar on this ability than they are different.

Why do our findings diverge from what might be thought of as conventional wisdom, i.e., that there is an overall sex difference in emotion recognition? One possible explanation is that of publication bias in this field. This account is supported by a recent meta-analysis of sex differences in emotion recognition ability that reported evidence for an excess of significant findings in the literature (Thompson & Voyer, 2014). For the field to move towards a consensus state, this suggests a need for strongly powered confirmatory studies with pre-registered experimental protocols.

What might account for this modest female advantage for recognising facial disgust? An interesting, if at present speculative, perspective posited to account for this more general sex difference concerns the unique selection pressures faced by females, including immunosuppression during pregnancy and over the menstrual cycle, higher risk of contracting sexually-transmitted diseases and transferring them to their offspring, and higher parental investment in infant protection (Fleischman, 2014). Given the greater vulnerability of females and their altricial offspring to contamination and infection, an evolutionary functionalist account theorises that females need to be more sensitive to cues of disgust (Curtis, Aunger & Rabie, 2004), which may include facial expressions.

In addition, Fleischman (2014) also hypothesised that males may show less disgust sensitivity than females due to a selection pressure to emphasise their robust immunity by displaying a relative indifference towards signals of disgust. Given that males consistently show greater risk-taking behaviour than females, Fessler, Pillsworth and Flamson (2004) suggested that this propensity may also extend to a higher willingness to approach sources of contamination in comparison to their female counterparts. By employing the minimum possible level of disease avoidance, it is suggested that males are highlighting their successful immune system and high genetic quality to potential reproductive mates.

It is conceivable, then, that such fitness imperatives shape the ability to recognise disgust in conspecifics in different ways across the sexes. The evolutionary adaptationist theory that disgust sensitivity may be functionally related to successful mating and reproduction is supported by the finding that sex differences are observed from puberty and young adulthood onwards, but no sex differences in disgust sensitivity emerge in child participants (Stevenson et al., 2010), as well as the finding that disgust sensitivity decreases across the lifespan as reproductive potential declines (Curtis et al., 2004). That said, we must reiterate that our study found largely overlapping distributions of ability to recognise disgust across women and men, so any evolutionary influences do not create major differences in this respect. Moreover, it is easy to think of ways in which cultural socialisation might tend to make women more sensitive than men to the importance of hygiene and risk of disease.

Some limitations of our studies require mention. Firstly, while MTurk samples are more diverse than student samples (Paolacci et al., 2010), they clearly do not form a representative sample of the population; for example, there was a greater proportion of female than male respondents in all three of our studies. And although our samples were ethnically diverse, they were mainly White and comprised solely of US residents; as such, all participants have been strongly exposed to Western culture. Further studies involving non-Western populations will therefore also be valuable in order to assess the robustness of the sex differences reported here. Secondly, here we focused on five basic emotions; however, emotional expressions are of course not restricted to these categories. Accordingly, work that explores a broader selection of emotions may reveal additional sex differences. Thirdly, while our stimuli are among the most carefully developed and validated for laboratory-oriented research, they clearly have limits with regards to ecological validity. As such, work that can take advantage of more naturalistic stimuli will be of value in future studies. Finally, we note that our bodily disgust recognition rate was low (in fact close to chance level) for both sexes, and this high level of noise may have led to a false negative result. Future work is recommended to address this issue.

**Conclusions**

In summary, across three independent samples we observed that females are superior to males in facial disgust recognition. This result is of particular note as this sex difference was observed across two very different sets of facial stimuli (static and dynamic). However, these group differences were modest in magnitude, and the overlap in ability between the two populations is substantial. No consistent evidence for further sex differences in emotion recognition ability was observed.

**References**

Alaerts, K., Nackaerts, E., Meyns, P., Swinnen, S. P., & Wenderoth, N. (2011). Action and emotion recognition from point light displays: An investigation of gender differences. *PloS One, 6*(6), e20989.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception, 33*(6), 717-746.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science, 338*(6111), 1225-1229.

Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: a comparison of self-report and performance measures of emotional intelligence. *Journal of Personality and Social Psychology, 91*(4), 780-795.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305-327.

Bruce, V., & Young, A. (2012). *Face perception*. Hove, East Sussex: Psychology Press.

Buck, R., Miller, R. E., & Caul, W. F. (1974). Sex, personality, and physiological variables in the communication of affect via facial expression. *Journal of Personality and Social Psychology, 30*(4), 587-596.

Calder, A.J., Young, A.W., Rowland, D., & Perrett, D.I. (1997). Computer-enhanced emotion in facial expressions. *Proceedings of the Royal Society of London, B. 264*(1383), 919-925.

Curtis, V., Aunger, R., & Rabie, T. (2004). Evidence that disgust evolved to protect from risk of disease. *Proceedings of the Royal Society of London, B. 271*, 131-133.

Darwin, C. (1872/1965). *The expressions of the emotions in man and animals.* New York: Oxford University Press.

de Gelder, B. & Van den Stock, J. (2011). The Bodily Expressive Action Stimulus Test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions. *Frontiers in Psychology, 2*(181), doi:10.3389/fpsyg.2011.0018.

Duesenberg, M., Weber, J., Schulze, L., Schaeuffele, C., Roepke, S., Hellmann-Regen, J., ... & Wingenfeld, K. (2016). Does cortisol modulate emotion recognition and empathy? *Psychoneuroendocrinology, 66*, 221-227.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124-129.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, California: Consulting Psychologists Press.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Fessler, D. M. T., Pillsworth, E. G., & Flamson, T. J. (2004). Angry men and disgusted women: An evolutionary approach to the influence of emotions on risk taking. *Organizational Behavior and Human Decision Processes, 95*(1), 107–123.

Fleischman, D. S. (2014). Women's disgust adaptations. In V.A. Weekes-Shackelford & T.K. Shackelford (Eds.), *Evolutionary perspectives on human sexual psychology and behavior* (pp. 277-296). New York: Springer.

Grimshaw, G. M., Bulman-Fleming, M. B., & Ngo, C. (2004). A signal-detection analysis of sex differences in the perception of emotional faces. *Brain and Cognition*, *54*(3), 248-250.

Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences*, *16*(5), 701-713.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, *85*(4), 845-857.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*(2), 201-206.

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behavior, 27*(6), 401-416.

Harmer, C. J., Goodwin, G. M., & Cowen, P. J. (2009). Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. *British Journal of Psychiatry, 195*(2), 102-108.

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta Psychologica*, *135*(3), 278-283.

Ioannidis, J. P., & Trikalinos, T. A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal*, *176*(8), 1091-1096.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201-211.

Kirouac, G., & Dore, F. Y. (1985). Accuracy of the judgment of facial expression of emotions as a function of sex and level of education. *Journal of Nonverbal Behavior*, *9*(1), 3-7.

Kret, M. E., & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia*, *50*(7), 1211-1221.

Krüger, S., Sokolov, A. N., Enck, P., Krägeloh-Mann, I., & Pavlova, M. A. (2013). Emotion through locomotion: gender impact. *PloS One*, *8*(11), e81716.

Lau, J. Y., Burt, M., Leibenluft, E., Pine, D. S., Rijsdijk, F., Shiffrin, N., & Eley, T. C. (2009). Individual differences in children's facial expression recognition ability: the role of nature and nurture. *Developmental Neuropsychology*, *34*(1), 37-51.

Lee, N. C., Krabbendam, L., White, T. P., Meeter, M., Banaschewski, T., Barker, G. J., ... & Frouin, V. (2013). Do you see what I see? Sex differences in the discrimination of facial emotions during adolescence. *Emotion*, *13*(6), 1030-1040.

Lewis, G. J., Lefevre, C. E., & Young, A. W. (2016). Functional architecture of visual emotion recognition ability: A latent variable approach. *Journal of Experimental Psychology: General*, *145*(5), 589-602.

Lyusin, D., & Ovsyannikova, V. (2016). Measuring two aspects of emotion recognition ability: Accuracy vs. sensitivity. *Learning and Individual Differences, 52,* 129-136.

Martinez, L., Falvello, V. B., Aviezer, H., & Todorov, A. (2016). Contributions of facial expressions and body language to the rapid perception of dynamic emotions. *Cognition and Emotion, 30*(5), 939-952.

Matsumoto, D., & Hwang, H. S. (2011). Judgments of facial expressions of emotion in profile. *Emotion*, *11*(5), 1223-1229.

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin, 126*(3), 424-453.

Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease-avoidance mechanism. *Psychological Bulletin, 135*(2), 303-321.

Olatunji, B. O., Williams, N. L., Tolin, D. F., Abramowitz, J. S., Sawchuk, C. N., Lohr, J. M., & Elwood, L. S. (2007). The Disgust Scale: item analysis, factor structure, and suggestions for refinement. *Psychological Assessment*, *19*(3), 281-297.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), doi: 10.1126/science.aac4716.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running Experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*(5), 411-419.

Rahman, Q., Wilson, G. D., & Abrahams, S. (2004). Sex, sexual orientation, and identification of positive and negative facial affect. *Brain and Cognition*, *54*(3), 179-185.

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior*, *12*(2), 139-148.

Rozin, P., Haidt, J., & McCauley, C. R. (1993). Disgust. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 575-594). New York: Guilford.

Rozin, P., Lowery, L., & Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. *Journal of Personality and Social Psychology*, *66*(5), 870-881.

Sawada, R., Sato, W., Kochiyama, T., Uono, S., Kubota, Y., Yoshimura, S., & Toichi, M. (2014). Sex differences in the rapid detection of emotional facial expressions. *PloS One*, *9*(4), e94747.

Sogon, S., & Izard, C. E. (1987). Sex differences in emotion recognition by observing body movements: A case of American students. *Japanese Psychological Research. 29*(2), 89-93.

Sokolov, A. A., Krüger, S., Enck, P., Krägeloh-Mann, I., & Pavlova, M. A. (2011). Gender affects body language reading. *Frontiers in Psychology*, *2*(16), 1-6.

Stevenson, R. J., Oaten, M. J., Case, T. I., Repacholi, B. M., & Wagland, P. (2010). Children’s response to adult disgust elicitors: Development and acquisition. *Developmental Psychology*, *46*(1), 165-177.

Surguladze, S. A., Young, A. W., Senior, C., Brébion, G., Travis, M. J., & Phillips, M. L. (2004). Recognition accuracy and response bias to happy and sad facial expressions in patients with major depression. *Neuropsychology, 18*(2), 212-218.

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion*, *28*(7), 1164-1195.

Tiddeman, B., Burt, D. M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *Computer Graphics and Applications, IEEE, 21*(5), 42–50.

Tybur, J. M., Lieberman, D., & Griskevicius, V. (2009). Microbes, mating, and morality: individual differences in three functional domains of disgust. *Journal of Personality and Social Psychology, 97*(1), 103-122.

Williams, M. A., & Mattingley, J. B. (2006). Do angry men get noticed? *Current Biology*, *16*(11), 402-404.

Young, A. (2016). *Facial expression recognition*. Hove, East Sussex: Psychology Press.

Young, A. W. (2018). Faces, people and the brain: the 45th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, doi:10.1177/1747021817740275.

Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology, 102*(4), 959-974.

Young, A., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion – stimuli and tests (FEEST).* Bury St Edmunds, England: Thames Valley Test Company.

**Figures**

Figure 1: Examples of each of the stimuli used: (A) static facial expressions, (B) frames from the dynamic body point-light displays, (C) a frame from the dynamic facial expressions, and (D) static body expressions. Note: Stimulus sets A and B were used in Studies 1, 2 and 3; Stimulus sets C and D were only presented in Study 2.

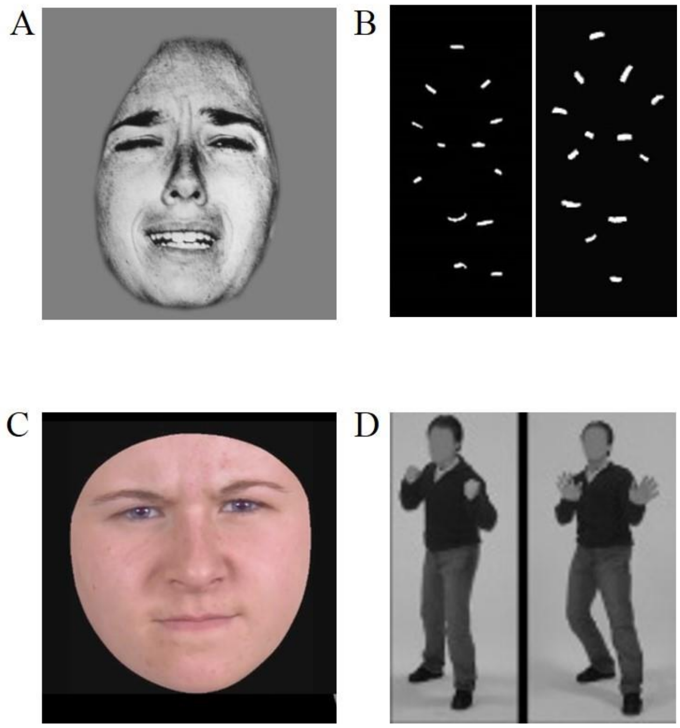
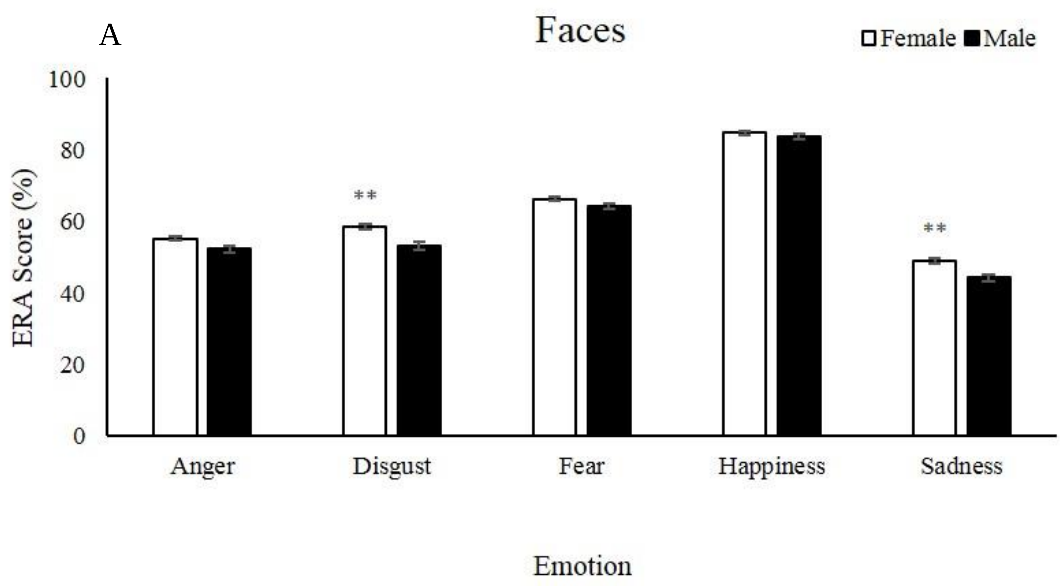


Figure 2: Mean emotion recognition accuracy in response to facial (A) and bodily (B) expressions for the five basic emotions in females and males in Study 1 (N = 1022 participants). Error bars represent standard error, and asterisks above the bars represent the results of the corrected *t* tests, where * indicates p<0.05 and ** indicates p<0.001.



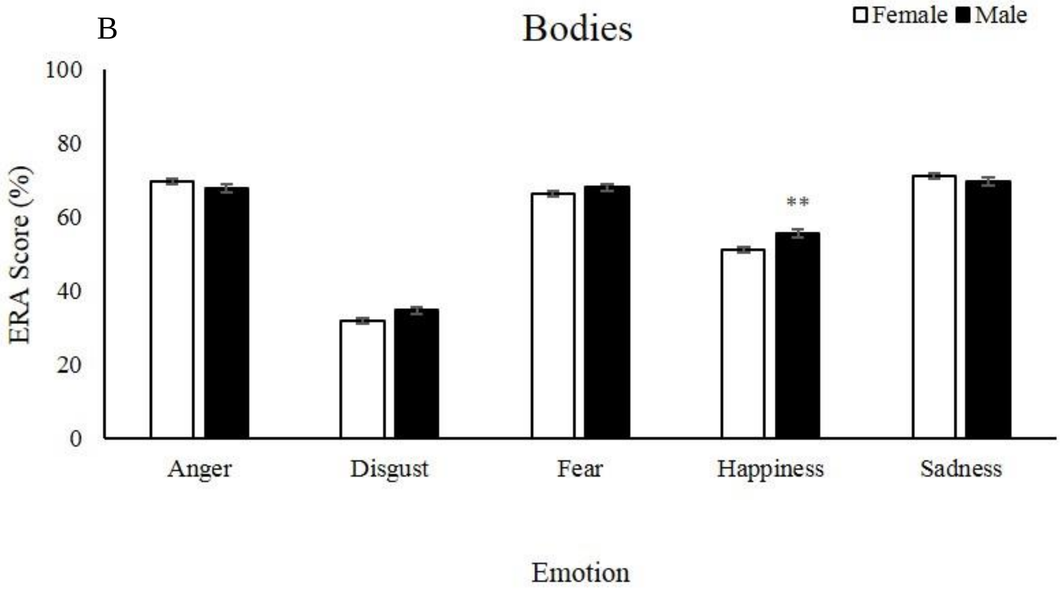



Figure 3: Mean emotion recognition accuracy in females and males in response to dynamic and static facial and bodily expressions for the three basic emotions that reached significance in Study 2 (N = 303 participants). Error bars represent standard error, and asterisks above the bars represent the results of the *t* tests, where * indicates p<0.05 and ** indicates p<0.001.

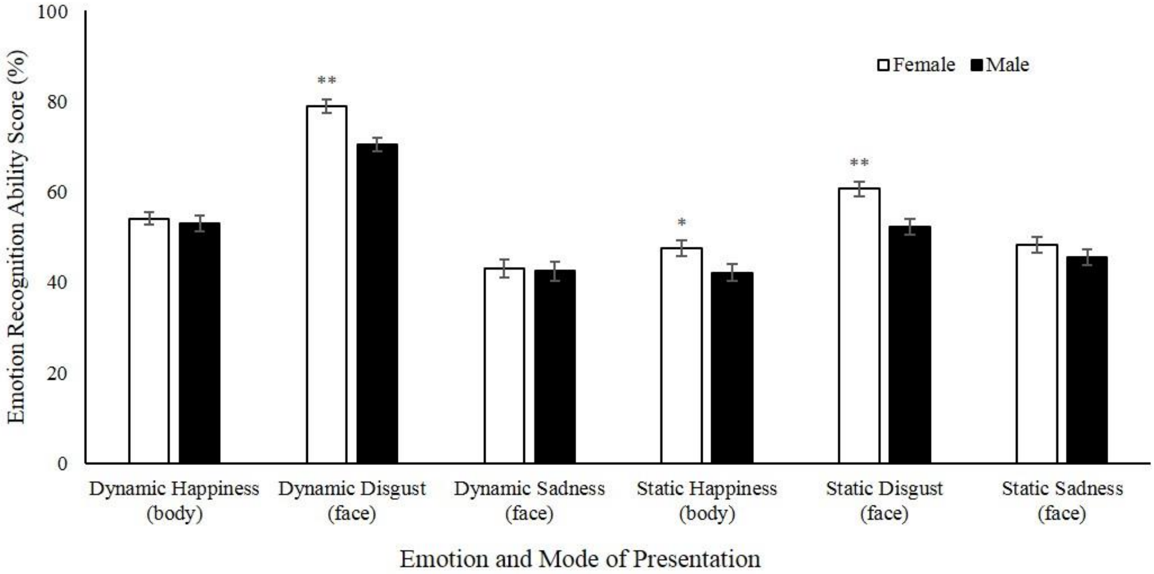


Figure 4: Mean emotion recognition accuracy in females and males in response to facial and bodily expressions for the three basic emotions that reached significance in Study 3 (N = 634 participants). Error bars represent standard error, and asterisks above the bars represent the results of the *t* tests, where * indicates $p<0.05$.