Computers in Human Behavior Reports 10 (2023) 100283

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

How do people respond to computer-generated versus human faces? A systematic review and meta-analyses

Elizabeth J. Miller [a], Yong Zhi Foo [a,b], Paige Mewton [a], Amy Dawel [a,*]

[a] School of Medicine and Psychology, The Australian National University, Canberra, ACT, 2600, Australia
[b] School of Biological Sciences, University of Western Australia, Crawley, WA, 6009, Australia

ARTICLE INFO

Keywords:
Virtual
Avatar
Trustworthiness
Generative adversarial network

ABSTRACT

Computer-generated (CG) beings are rapidly infiltrating the human social world. Yet evidence about how humans respond to CG faces is mixed. The present systematic review and meta-analyses aimed to synthesise empirical evidence from studies comparing people's responses to CG and human faces, across key face processing domains of interest to psychology, neuroscience, and computer science. We tested whether effects were moderated by the perceived realism of CG relative to human faces, and whether CG and human faces showed the same identity or not. We hypothesised that people would be able to tell CG and human faces apart, and that other types of responses would favour human over CG faces. While results supported our hypotheses across several domains (perceptions of human-likeness, face memory, first impressions, emotion labelling), some responses did not differ for CG and human faces (quality of interactions, emotion ratings, facial mimicry, looking behaviour). We also found a reduced inversion effect for CG relative to human faces, though only minimal data were available for hallmark face effects (ORE, N170 and FFA responses). Overall, findings highlight potential strengths and challenges of using CG faces across a range of applications, including e-health, social companionship, video-gaming, and scientific work.

Our social world is rapidly expanding to include computer-generated (CG) beings. CG beings are taking on important roles that have traditionally been occupied by humans, in customer service (Wang & Fodness, 2010), psychological therapy (Cooper et al., 2019; Dellazizzo et al., 2018), medical training (Andrade et al., 2010; Guise et al., 2012), social companionship (e.g., as carers and dating partners, Abbott & Shaw, 2016; Craft, 2012), and in social media. For instance, Lil Miquela (https://www.instagram.com/lilmiquela/?hl=en), a CG model and musician, has more than 3 million Instagram followers and is estimated to generate an annual $12 million USD by advertising for popular brands such as Calvin Klein, Samsung, and Prada (https://businessesgrow.com/2020/11/25/lil-miquela/). This shift in our social landscape highlights the importance of understanding how humans experience CG beings.

The present systematic review and meta-analyses provides a critical first synthesis of the scientific research comparing people's responses to CG and human faces, integrating findings across disparate disciplines, including psychology, neuroscience, and computer science. The now considerable literature on this topic has produced mixed findings—with some showing people respond similarly to CG and human faces (e.g., Javor et al., 2016; Joyal et al., 2014) but others suggesting important differences (Balas & Pacella, 2015; Crookes et al., 2015; Kätsyri, 2018)—warranting this synthesis. We focussed on faces because of the central role they have in human social interaction, and high levels of scientific interest in how they are processed. For example, faces provide critical information about identity and emotion (Calder & Young, 2005), and first impressions of them are so influential that people with trustworthy-looking faces are more likely to be elected as politicians and receive less severe criminal sentences, irrespective of the person's actual trustworthiness (Todorov et al., 2008). Scientific research has also identified behavioural (ORE: Meissner & Brigham, 2001; part-whole effect: Tanaka & Farah, 1993; inversion effect: Yin, 1969; composite effect: Young et al., 2013) and neural markers (Bentin et al., 1996; Kanwisher et al., 1997) that are unique to faces, highlighting their importance in human social behaviour.

Why are CG beings so popular? In the long run, CG beings may be cheaper to "employ" than humans, and can provide consistent service, unaffected by mood, illness, or personal circumstances as a human might be. CG beings can be tailored to a role or a person's preferences.

* Corresponding author. School of Medicine and Psychology, The Australian National University, Canberra, ACT, 2600, Australia.
E-mail addresses: elizabeth.miller@anu.edu.au (E.J. Miller), fooyongzhi@gmail.com (Y.Z. Foo), paige.mewton@anu.edu.au (P. Mewton), amy.dawel@anu.edu.au (A. Dawel).

https://doi.org/10.1016/j.chbr.2023.100283
Received 18 September 2022; Received in revised form 24 January 2023; Accepted 30 March 2023
Available online 7 April 2023

For instance, CG therapists can be made to appear trustworthy, or a CG dating partner tailored to a person's attractiveness preferences. Some people may prefer the anonymity that CG beings offer (Lucas et al., 2014, 2017). CG faces are popular with scientists too; they are easy to generate, and allow scientists to systematically manipulate variables of interest more efficiently than photographs of humans allow (e.g., activation of a facial expressions; Krumhuber et al., 2012). As a result, CG faces are increasingly being used as proxies for human ones in research (Dawel et al., 2021).

Here, we systematically reviewed literature that has investigated people's responses to CG compared to human faces, testing for overall answers using meta-analyses. If people's responses to CG and human faces are the same, CG faces *may* successfully perform some human social functions. For instance, if CG beings elicit as much trust as humans, virtual influencers like Lil Miquela may be just as effective on Instagram as human brand ambassadors. Alternatively, if responses to CG and human faces differ, CG faces may reduce the quality of social interactions. For example, CG faces may not adequately mimic human eye-contact or animacy (Kätsyri et al., 2020; Syrjämäki et al., 2020), which may reduce the quality of rapport with CG beings and, in turn, social and other outcomes (e.g., reducing therapeutic effects in mental health treatment). In science, similar responses to CG and human faces might suggest researchers can save the time and cost associated with developing human face databases, and use CG ones in empirical work instead. However, different responses to CG and human faces would suggest that CG faces are not adequately tapping into human face processing systems as researchers intend.

Our systematic review and associated meta-analyses addressed core questions such as: Can people tell CG faces apart from human ones? Is identity or expression recognition poorer for CG than human faces? And do people evaluate CG faces differently to human ones (e.g., perceptions of trustworthiness)? We also reviewed hallmark effects that have been of longstanding interest in face perception research, such as the inversion (turning a face upside-down makes it disproportionately more difficult to recognise: Yin, 1969) and other race effects (worse recognition of faces of another race compared to one's own; Meissner & Brigham, 2001), and neural markers of face processing such as the N170 ERP (Bentin et al., 1996). Our general hypothesis was that, because CG faces lack some of the key features of human ones (e.g., reduced animacy in eyes & overly smooth skin: Vaitonytė et al., 2021; Balas & Tonsager, 2014), people's responses would be poorer or reduced for CG compared to human faces. The results of the present study are presented as meta-analyses for any domain where there was sufficient suitable data, and as narrative reviews for domains where data were not suitable for meta-analysis (e.g., because there were insufficient data or methodological differences between studies). The following introduction also provides key background for those domains where it was possible to perform meta-analyses.

Can people tell CG faces apart from human ones?

Most of the evidence from individual studies indicates people can reliably tell CG faces apart from human ones (Bailenson et al., 2004; Carlson et al., 2012; Cheetham, 2011; Cheetham et al., 2014; Fan et al., 2012; Gonzalez-Franco et al., 2016; Kätsyri, 2018). However, much of this evidence is from older CG faces that are obviously CG (see Fig. 1). People have more difficulty telling the latest hyper-realistic CG faces apart from human ones (Nightingale & Farid, 2022). Our meta-analyses compared people's ability to tell CG and human faces apart relative to chance (i.e., when choosing whether a face is CG or human, chance is 50%). We also tested the role of important moderators, including how realistic CG faces are relative to their human counterparts. We predicted effects would be larger (i.e., greater accuracy relative to chance) when there are larger differences in the perceived realism of the CG and human faces, making it easier to tell them apart. We hypothesised that:

H1. People can reliably tell CG and human faces apart, and are better at doing so when there is a greater difference in their perceived realism.

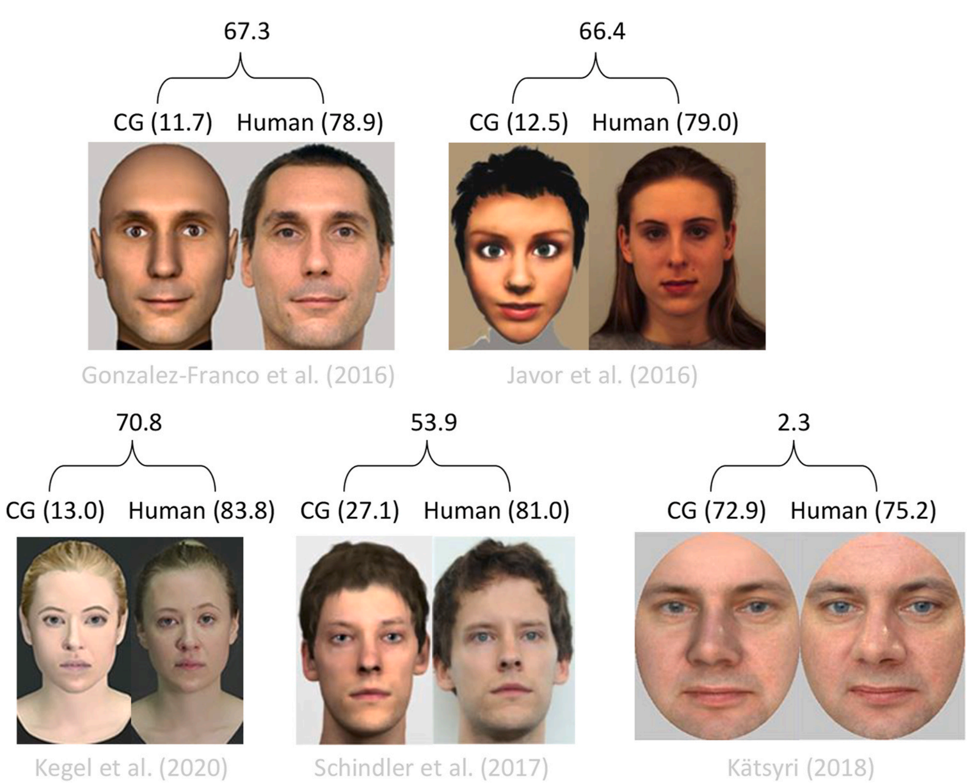

Fig. 1. Examples of CG and human face stimuli from papers included in the meta-analysis.
Note. Numbers refer to mean realism ratings (see Supplement S3) on a scale of 0 (extremely unrealistic) to 100 (extremely realistic). All images presented here are covered by Creative Commons (CC-BY).

Are CG faces perceived as less human-like?

A related question concerns the extent to which CG faces are perceived as being human-like. Many studies have asked questions like, "How human-like/realistic is this face?" (e.g., Kätsyri, 2018; MacDorman & Chattopadhyay, 2017) or "How alive is this face?" (e.g., Krumhuber et al., 2015). Also of interest are factors related to the uncanny valley hypothesis (Mori, 1970), whereby CG faces that are very human-like, but not quite human, are perceived as more strange and eerie than CG faces which are obviously non-human (for meta-analysis, see Diel et al., 2021). CG skin and eyes appear clearly different to human faces (Balas & Tonsager, 2014; Vaitonytė et al., 2021), cueing people to their non-humanness. Accordingly, individual studies have found CG faces are perceived as less human-like (Bartneck et al., 2007; Kätsyri, 2018; Kätsyri et al., 2019; Patel & MacDorman, 2015; Tinwell et al., 2011, 2015; Tinwell & Sloan, 2014), realistic (Dai & MacDorman, 2018; Gaither et al., 2019; Gonzalez-Franco et al., 2016; Kätsyri et al., 2020), lifelike (Carter et al., 2013), and animated (Krumhuber et al., 2015), compared to human faces. CG faces are also rated as more uncanny (Tinwell et al., 2013), strange (Carter et al., 2013; Cheetham et al., 2014, 2015; Tinwell et al., 2011, 2015; Tinwell & Sloan, 2014), eerie (Chattopadhyay & MacDorman, 2016; Cheetham et al., 2015; Kätsyri, 2018; Kätsyri et al., 2019; MacDorman & Chattopadhyay, 2017; Patel & MacDorman, 2015), and creepy (Carter et al., 2013) than human faces. To the best of our knowledge, only one study (Dai & MacDorman, 2018) has found no significant difference in eeriness ratings for CG and human faces. We hypothesised that:

H2. People perceive CG faces as less human-like than human faces, and this difference is larger when there is a greater difference in their perceived realism.

Is memory for CG faces poorer than for human faces?

Face memory studies mostly used old/new recognition tasks, where participants learn one set of faces and then judge whether they are old or new when mixed with distractor faces. Most individual studies have found people are worse at remembering CG than human faces (Bailenson et al., 2003, Bailenson et al., 2004; Balas & Pacella, 2015; Carlson et al., 2012; Crookes et al., 2015; Kätsyri, 2018; Seo et al., 2017; but also cf. Carlson et al., 2012; Kätsyri, 2018; Seo et al., 2017 for individual experiments finding no significant difference). There is a strong theoretical basis for expecting this difference in face space theory (Valentine, 1991; Valentine et al., 2016). Face space proposes people's face perception is finely tuned to the information that is important for recognising the faces they encounter while this system is developing, primarily during childhood (McKone et al., 2019; Singh et al., 2021). This theoretical account suggests two reasons why face memory might be impaired for CG faces. First, if study participants were not exposed to CG faces during the period their face space developed, their face space may not be optimally tuned for CG face perception. Second, CG faces lack variation in some of the information that human-tuned face space probably uses (e.g., reflectance/skin texture; Itz et al., 2014) which may reduce CG face perception sensitivity. We hypothesised that:

H3. People's memory is poorer for CG than human faces.

Is the quality of interactions with CG faces impoverished?

There is a growing body of evidence concerning people's experiences, or expected experiences, of interactions with CG faces. Factors such as how much empathy people feel or how much they enjoy interacting with CG faces are of interest because they might affect the quality of human-CG social connections, and the roles that CG beings could successfully fulfill. For example, rapport is central to human-human therapy (Flückiger et al., 2018) and high-quality social connections increase human wellbeing (Cruwys et al., 2014). While several studies have found interactions are impoverished for CG compared to human faces (Chattopadhyay & MacDorman, 2016; Cheetham et al., 2015; Hyde et al., 2014; Seo et al., 2017), some evidence favours CG over human faces (Dai & MacDorman, 2018; MacDorman, 2019; Roth et al., 2019). For example, people are more willing to follow the advice of a CG than a human doctor, and may enjoy interacting with them more (Dai & MacDorman, 2018; MacDorman, 2019). Findings like these may reflect the benefits of anonymity in specific settings (e.g., sharing personal health information). For the purposes of the present review, we followed our general hypothesis and predicted that:

H4. People's interactional experiences will be impoverished for CG relative to human faces.

Do first impressions differ for CG and human faces?

The initial impressions people form about faces are of considerable interest across the face literature (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Vernon et al., 2014). Observers judge whether a person is trustworthy or not within 100 ms of viewing their face (Willis & Todorov, 2006). Despite being only moderately accurate (Foo et al., 2021), these judgements have critical real-world impacts, including influencing election outcomes (Ballew & Todorov, 2007; Little et al., 2007; Todorov, Mandisodza, Goren, & Hall, 2005), criminal sentencing decisions (Wilson & Rule, 2015), and success on Airbnb (Ert et al., 2016). It is unsurprising then that people's first impressions of CG faces have received significant attention. The most common method used to investigate first impressions is for observers to rate a particular trait, such as how trustworthy, warm, or competent a person is, viewing only their face. While several studies have found first impressions are less favourable for CG than human faces (for trustworthiness: Balas & Pacella, 2017; Gong, 2008; attractiveness: Patel & MacDorman, 2015; friendliness/likeability: Bartneck et al., 2007; Gong, 2008; Tinwell & Sloan, 2014; competence: Gong, 2008; warmth: Chattopadhyay & MacDorman, 2016; MacDorman & Chattopadhyay, 2016), many others have found CG faces are evaluated just as favourably as human faces on many of the same traits (for trustworthiness: Javor et al., 2016; Riedl et al., 2014; Patel & MacDorman, 2015; attractiveness: Carlson et al., 2012; Cheetham et al., 2015; likeability: Carter et al., 2013; warmth: Dai & MacDorman; MacDorman & Chattopadhyay, 2017; competence: Dai & MacDorman, 2018; MacDorman & Chattopadhyay, 2017; Patel & MacDorman, 2015; goodwill: Patel & MacDorman, 2015). However, new evidence suggests that people trust hyper-realistic AI-generated faces *more* than human ones (Nightingale & Farid, 2022). Given this mixed evidence, our specific hypothesis followed our general hypothesis:

H5. People's first impressions are less favourable for CG than human faces.

Responses to emotional facial expressions

People's responses to emotional facial expressions may be impoverished for CG relative to human faces because CG faces lack some of the fine-grained detail (e.g., skin wrinkles; Miller et al., 2020) and physiological cues (e.g., tears: Balsters et al., 2013; pupil changes: Demos et al., 2008; colour: Peromaa & Olkkonen, 2019) that assist perception of emotion in human faces. Research investigating emotion perception for CG relative to human faces has focused on Ekman's original six basic emotions (i.e., anger, disgust, fear, happiness, sadness, and surprise; Ekman, 1992) using three different methods: emotion labelling, emotion ratings, and facial mimicry measured with facial electromyography (fEMG). In emotion labelling tasks, participants choose what emotion a face shows, usually from a list (e.g., happy, sad, etc.). In emotion rating tasks, participants may rate how intense, arousing, or positive-negative an expression is. Facial mimicry measures the extent to which the observer's facial muscles are activated in response to a facial expression (Dimberg, 1982).

Overall, emotion labelling tends to be poorer for CG than human faces (Afzal et al., 2009; Baltrušaitis et al., 2010; Costantini et al., 2004, 2005, Fabri et al., 2004; Kätsyri et al., 2003; Moser et al., 2007; but cf. Bartneck, 2001), and CG facial expressions are often rated lower on intensity, arousal, and valence than human ones (Craig et al., 2012; Kegel et al., 2020; Mühlberger et al., 2009; but cf. Jackson et al., 2015; Kätsyri et al., 2020). There is also some evidence that facial mimicry is weaker for CG than human faces (Philip, Martin & Clavel, 2018; but cf. Joyal et al., 2014 for no difference in mimcry). However, labelling accuracy is occasionally better for CG than human faces for some individual emotions (e.g., for anger, Kätsyri et al., 2020; fear, Dyck et al., 2008, Dyck et al., 2010; Kätsyri et al., 2020; Milcent et al., 2019; and sadness, Dyck et al., 2008, Dyck et al., 2010; Milcent et al., 2019). One important factor that may moderate responses to facial expressions is whether stimuli are static (still) or dynamic (moving). While most studies use static facial expressions (Dawel et al., 2021), real life facial expressions are dynamic, and there is considerable evidence that human perception advantages dynamic expressions over static ones (Arsalidou et al., 2011; Biele & Grabowska, 2006; Calvo et al., 2016; Recio et al., 2011). Thus, any differences between CG and human faces may be exaggerated for dynamic stimuli. We therefore included emotion category (e.g., anger, disgust, etc.) and static/dynamic status as moderating variables in our expression meta-analyses. Overall, we hypothesised that:

H6. People's responses to facial expressions of emotion are impoverished for CG relative to human faces (e.g., reduced labelling accuracy, weaker emotion-related ratings, weaker facial mimicry).

Do people look at CG and human faces differently?

Researchers have primarily used eye tracking to investigate how often and for how long people look at CG and human faces. More frequent or longer looking at CG faces may signal greater interest but may equally imply a greater sense of eeriness. While people seem to look just as often at CG and human faces (Cheetham et al., 2013; Chen et al., 2019) and face features (e.g., eyes or mouth; Cheetham et al., 2013; Zhou et al., 2016), the evidence regarding looking time is mixed. Some studies have found people look at CG faces and face parts for less time than human ones (e.g., Cheetham et al., 2013; Joyal et al., 2014; Lewkowicz & Ghazanfar, 2012) but others have found the opposite pattern (Lewkowicz & Ghazanfar, 2012; Ni et al., 2018). Because it was difficult to predict whether looking frequency and duration should be reduced because CG faces do not attract the same attention as human ones, or increased because CG faces are unusual or eerie, we had no specific hypothesis about the direction of effect for looking behaviour. We simply tested *if* there are any differences in looking behaviour for CG compared to human faces.

Hallmark effects associated with perception of human faces

Human face processing is characterised by several hallmark effects that are theorised to reflect the unique processing of human faces (e.g., holistic processing; Valentine, 1991; Young et al., 2013; see McKone et al., 2007 for a review). Here, we focussed on hallmark effects for which there is enough data suitable for meta-analysis comparing CG with human faces: the inversion effect (Yin, 1969), the other race effect (Meissner & Brigham, 2001), and the N170 ERP component (Bentin et al., 1996).[1] Comparing these hallmark effects in CG and human faces may help us understand why memory and other responses are potentially impoverished for CG faces. Our first question in this section was whether these effects are evident for CG faces at all. Our second question was, if these effects are evident, whether they are the same size for CG and human faces.

Inversion effect. Relative to other objects, turning a face upside-down makes it disproportionately more difficult to recognise (Rossion, 2008; Yin, 1969). Some studies have found evidence of an inversion effect for CG faces (Balas & Pacella, 2015; Kätsyri, 2018; Matheson et al., 2012; Matheson & McMullen, 2011) but others have not (e.g., Experiments 1 & 2 in Carlson et al., 2012). Although the one study that compared the size of the inversion effect for CG and human faces found there was no significant difference in the size of the effect (Balas & Pacella, 2015), this may be due to the small sample size ($N = 12$). Following our general hypothesis, our specific hypothesis for this domain was:

H7. CG faces may produce an inversion effect, but its size will be reduced relative to human faces.

Other-race effect (ORE). People tend to be better at recognising faces of their own races than faces of another race (McKone et al., 2019; Meissner & Brigham, 2001). While CG faces seem to produce an ORE (Balas et al., 2011; Crookes et al., 2015; Hourihan et al., 2013; Matheson & McMullen, 2011; Papesh & Goldinger, 2009), it is unclear whether the effect is as strong as for human faces (e.g., Crookes et al., 2015). We hypothesised that:

H8. CG faces may produce an other-race effect, but its size is reduced relative to that for human faces.

N170. The N170 ERP is a negative waveform that appears approximately 170ms after a person sees a face, thought to reflect specialised neural processing for faces (Bentin et al., 1996; Rossion & Jacques, 2012). There is considerable evidence that CG faces elicit the N170 response (Geiger & Balas, 2021; Gonzalez-Franco et al., 2016; Kala et al., 2021; Mühlberger et al., 2009; Naples et al., 2015; Perizzolo Pointet et al., 2020; Schindler et al., 2017; Sollfrank et al., 2021; Wang, Xu, et al., 2019). However, it is unclear whether this response is the same for CG and human faces, with studies producing conflicting findings (Gonzalez-Franco et al., 2016; Mühlberger et al., 2009; Schindler et al., 2017; Sollfrank et al., 2021). We hypothesised that:

H9. CG faces may produce an N170 response, but its size is reduced relative to that for human faces.

The present study

The present study is the first systematic review of studies that compare people's responses to CG and human faces, drawing on literature from psychology, neuroscience, and computer science. Because the theoretical basis, empirical methods, and evidence differ across the different face processing domains, we present findings for each domain separately (Supplement S9). Where sufficient data were available within a domain, we quantitatively synthesised the evidence via meta-analyses. For all meta-analyses, we included as moderators: the perceived realism of CG relative to human faces; whether the CG and human stimuli showed the same identities (matched) or not (unmatched); and year of publication. We also included additional moderators in analyses where they are of particular interest to that domain (e.g., emotion category and static/dynamic format for the facial expressions domain). Our hypotheses for each domain are summarised in Table 1. Some domains could not be meta-analysed (e.g., because there were insufficient data or methodological differences between studies), in which case findings were synthesised by narrative review. Namely, for brain imaging and EEG studies testing components other than the N170; perceptions of face gender, sexuality, and race; and identification and pronunciation of speech.

[1] There are also some data comparing fusiform face area (FFA) responses to CG and human faces, but these were insufficient for meta-analysis so we provide a narrative review in the Results instead. Other important hallmark effects, including the composite (Young et al., 2013) and part-whole (Tanaka & Farah, 1993) effects, are yet to be tested in paradigms that directly compare responses to CG and human faces.

Table 1
Meta-analysis hypotheses for each face processing domain.

Face processing domain	Hypothesis	CG-human difference?
General (across domains)	H: People's responses are poorer or reduced for CG compared to human faces.	
Can people tell CG faces apart from human ones?	H1: People can reliably tell CG apart from human faces, and are better at doing so when there is a greater difference in their perceived realism.	Yes
Are CG faces perceived as less human-like?	H2: People perceive CG faces as less human-like than human faces, and this difference is larger when there is a greater difference in their perceived realism.	Yes
Is memory for CG faces poorer than for human faces?	H3: People's memory is poorer for CG than human faces.	Yes
Is the quality of interactions with CG faces impoverished?	H4: People's interactional experiences will be impoverished for CG relative to human faces.	No
Do first impressions differ for CG and human faces?	H5: People's first impressions are less favourable for CG than human faces.	Sometimes
Responses to emotional facial expressions	H6: People's responses to facial expressions of emotion are impoverished for CG relative to human faces (e.g., reduced labelling accuracy, weaker emotion-related ratings, weaker facial mimicry).	Sometimes
Do people look at CG and human faces differently?	Non-directional	No
Hallmark effects associated with perception of human faces		
Inversion effect	H7: CG faces may produce an inversion effect, but its size is reduced relative to that for human faces.	Yes
Other-race effect	H8: CG faces may produce an other-race effect, but its size is reduced relative to that for human faces.	No
N170	H9: CG faces may produce an N170 response, but its size is reduced relative to that for human faces.	No

Method

Search strategy

The three target databases—ACM Digital Library (computer science), PsycINFO, and Scopus (psychology and neuroscience)—were searched on 26 September 2019, and updated prior to data analysis on 26 February 2021. The searches combined three sets of terms: face terms (“emotion”, “expression”, “face”, “facial”, “gaze”, “identity”), CG terms (“agent”, “animat*”, “artificial*”, “avatar”, “character”, “characters”, comput*”, “FaceGen”, “FaceMaker”, “FACSGen”, “fake”, “GAN”, “generative adversarial network”, “photorealistic”, “poser”, “virtual”), and human terms (“human*”, “natural”, “photo*”, “real*”). We searched abstracts, titles, and keywords in PsycINFO and Scopus, and abstracts in ACM Digital Library (title and keyword searches are not available in ACM Digital Library). In PsycINFO and Scopus, we used proximity operators to restrict results to those where CG and human terms appeared within five words of “face” (proximity operators are not available in ACM Digital Library). In ACM Digital Library, the “comput*” term was replaced with “computer generated” to exclude large numbers of irrelevant results (e.g., articles about human-computer interfaces). Apart from these small differences, searches were consistent across databases. The full search strategy is reported in Supplement S1.

Eligibility criteria

To be included in the review, studies were required to:

- Measure a behavioural (e.g., accuracy, ratings of attributes such as trustworthiness) and/or neural (e.g., EEG ERPs) response in human participants, to both CG and human faces.
- Use the same experimental paradigm to measure the same behavioural and/or neural response in both CG and human faces (e.g., studies comparing responses to dynamic CG faces with responses for static human faces were excluded).
- Focus on responses to faces (e.g., studies comparing responses to full-bodied CG and human beings were excluded, unless they manipulated the face independently of the body) or face parts (e.g., responses to CG and human eyes).
- Use CG face stimuli created with software, including but not limited to FaceGen (Singular Inversions Inc., 2009), FACSGen (Krumhuber et al., 2012), Poser (Curious Labs, Santa Cruz, CA), GANs (Karras et al., 2021), avatar faces or real-time animation methods. Stimuli embedded within virtual reality were also included, provided other criteria were met. Faces that were mechanical (e.g., robot faces), cartoons, dolls, line drawings, or schematic faces were excluded.
- Use human face stimuli that are static photographs or videos of real human faces, including video chat methods (e.g., Skype or Zoom). Videos could include speech, if it was matched for the CG and human face conditions (e.g., Hetherington, 2015 did not meet this criterion). Studies in which the human stimulus was physically present (e.g., in the same room as participants) were excluded.
- Be published in a peer-reviewed journal or conference proceeding, with the full text available in English.

Eligibility screening

Fig. 2 illustrates the full eligibility screening process, conducted by EJM, including records retained and excluded at each stage. Exact duplicates were removed first, then all titles and abstracts were screened using Rayyan (Ouzzani et al., 2016). Full text articles were obtained and assessed against the eligibility criteria for all remaining records; the reasons for excluding any articles at this stage were documented (see Fig. 2). To identify other potentially relevant articles, the reference lists of all included articles and the Google Scholar profiles of authors with three or more included articles were hand-searched. EJM also e-mailed authors with three or more included articles asking them to identify any additional relevant published articles. Articles encountered in other ways by EJM or AD (e.g., shared by colleagues, recommended online during the article download process) which met eligibility criteria were also included.

Data coding and effect size calculation

All data was coded by EJM using Qualtrics (see Supplement S7 for data extraction form). We used Cohen's *d* (Cohen, 1988) as our effect size measure because it is commonly used and easy to interpret. We calculated Cohen's *d* using formulas provided in Nakagawa and Cuthill (2007), using means and SDs if they were available (extracted from

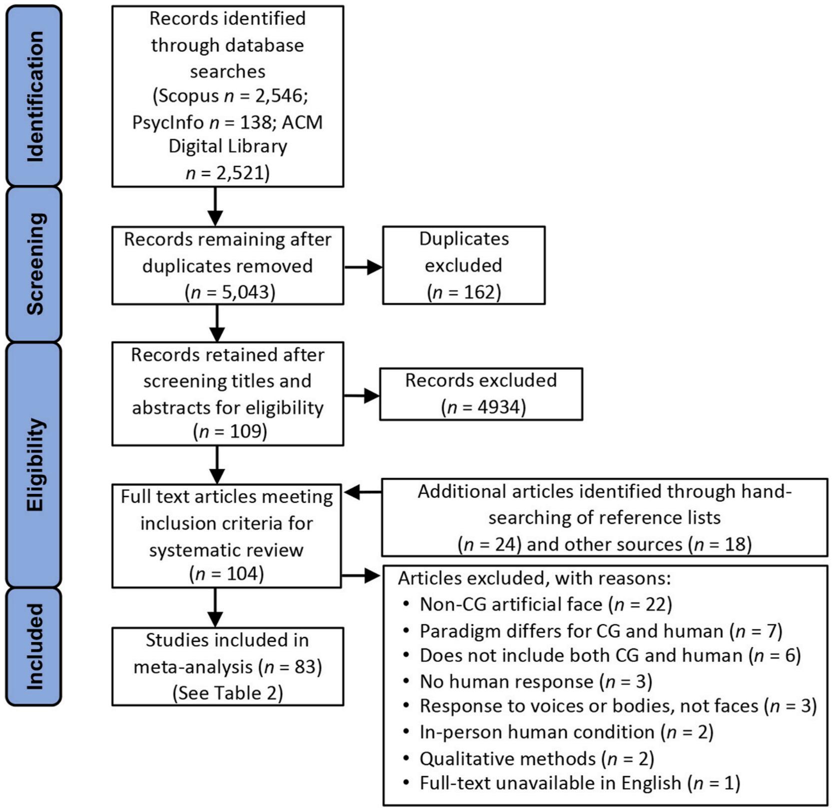

Fig. 2. PRISMA flow chart showing search and article screening process.

figures using WebPlotDigitiser [Rohatgi, 2021]) or converted using Excel (e.g., for SEM to SD) if required. Other information (e.g., *t*-values) was used only if means and SDs were unavailable.[2] If responses to CG and human faces were compared in multiple conditions (e.g., for different viewpoints or emotions) we calculated separate effect sizes for each condition if data were available. Otherwise, we calculated a single effect size for all of the conditions combined. For papers that used stimuli which morphed CG with human faces (e.g., Chattopadhyay & MacDorman, 2016), we included only the endpoints of the morph continua (i.e., 100% CG and 100% human) unless the only information available combined responses across morph levels (e.g., Cheetham et al., 2014). To avoid multiple comparisons, we retained responses to only the best quality or most realistic condition in the meta-analysis. For example, for Crookes et al. (2015), we included data from FaceGen stimuli generated from photographs of humans and not that from randomly generated FaceGen stimuli. Where papers provided insufficient information to calculate effect sizes, at least two attempts were made to obtain this information from the authors.

In addition to effect sizes, information extracted from each paper included participant demographic information (e.g., age, race, sex) and stimulus details (e.g., software used to create CG faces, source database for human faces, race and sex of stimuli).

[2] As per Nakagawa and Cuthill (2007), the formula to calculate Cohen's *d* from *t*-values and variance for within-subjects effect sizes requires imputation of the correlation between variables. Although most studies included in the meta-analysis used within-subjects designs, correlations were only available in relation to ~20% of effect sizes. We therefore used average correlation as a substitute, weighted by *N* of participants, from that face processing domain if available (e.g., the average weighted correlation from first impressions studies) or across the entire dataset (0.26) if none were provided for that domain.

Meta-analyses

To assess reliability, EJM and PM independently extracted the variables used in meta-analyses for 10 articles randomly selected from the N = 83 meta-analysis articles. This included all variables that contributed to effect size calculation and moderation effects (e.g., publication year, identity matched or non-matched, sample sizes, means and SEMs or SDs, plus *N* identities which contributed to Table 2). Average agreement between the two coders across the 10 articles was 85% ($SD = 19.6\%$; see Supplement S8 for details). All differences were resolved via discussion. Reasons for the differences included that EJM used additional data sent by the authors of the target article (Joyal et al., 2014) that PM had not accessed, and that data was inconsistently reported within an article (e.g., N reported in participants section differed from N reported in table of demographic information; Hernandez et al., 2009). There were only two instances in which a difference was due to a mistake made by EJM (1 = misunderstood N CG identities; 1 = typo), indicating that the coding of the remaining data by EJM was highly reliable.

We ran multilevel meta-analyses using mixed linear models (MLMs) with the Metafor package in R (Viechtbauer, 2010; See https://osf.io/fxcku/ for full data and R code). All models used the Restricted Maximum Likelihood (REML) method for estimating effect sizes and 95% confidence intervals. The general analysis procedure was the same for each face processing domain. First, we checked for heterogeneity amongst effect sizes. If heterogeneity was 0%, we did not proceed with moderator analysis. Next, we ran overall meta-analytic models. Sources of non-independence (multiple effects from the same paper and multiple effects from the same group of participants) were controlled for by entering them as random factors. For each model, we entered random factors (paper id, effect size id, participant group) first. Next, each moderator was added separately, provided the ratio of the smallest to largest number of cases in each level of categorical moderators was larger than 1:10 (Foo et al., 2017). All random effects and moderators

Table 2
Number of papers, effect sizes, participants and faces included in meta-analyses for each face processing domain.

Face processing domain	Meta-analysis (MA) or narrative only (N)?	N articles[a]	N effect sizes	N participants	N CG/human faces
Telling CG and human faces apart					
d'	MA	3	19	867	80/80
Accuracy	MA	10	91	2580	148/148
Dissimilarity ratings	MA	3	10	66	45/45
Human-likeness	MA	23	111	2365	162/166
Face memory					
Accuracy and d'	MA	11	46	824	463/407
False alarms	MA	5	12	413	320/256
RT	MA	2	11	326	77/77
Interactional experiences	MA	17	139	1445	73/73
First impressions	MA	26	164	3755	235/259
Responses to emotional facial expressions					
Emotion labelling	MA	13	80	549	77/104
Emotion ratings	MA	9	87	442	78/104
Facial mimicry	MA	2	14	71	10/10
Looking behaviour	MA	7	100	235	42/67
Hallmark effects					
Inversion effect	MA	4	22	283	150/150
Other-race effect	MA	2	14	167	192/192
N170	MA	3	5	155	14/74
Other EEG/ERP	N	–	–	–	–
Brain imaging	N	–	–	–	–
Speech	N	–	–	–	–
Gender, race and sexuality judgements	N	–	–	–	–

[a] Number of articles does not sum to the total number of articles included in the meta-analyses because articles were able to contribute effects across multiple face processing domains.

were then added to a full model to calculate omnibus moderator effects. We included our three primary moderators (difference in perceived realism for CG and human faces, identity matched/not matched, publication year) in all analyses but report results only where moderators were significant, or results differed across moderator levels.

Results

Analysis of where effects were drawn from highlighted the multidisciplinary interest in this topic, with significant interest from psychology, computer science and neuroscience, as illustrated in Fig. 3.

Heterogeneity among effect sizes was moderate to large across all face processing domains (I^2 range: 36.72–99.82%; See Supplement S4 for I^2 for individual domains), so we proceeded with moderator analyses with two exceptions. We were unable to include moderators for facial mimicry because there were only two papers, and we did not conduct moderator analysis for the hallmark face effects as our research questions differed from that of the other face processing domains.

Can people tell CG faces apart from human ones?

As hypothesised (**H1**), meta-analysis found strong evidence that people can tell CG faces apart from human ones. This conclusion was supported by three separate meta-analyses showing that: (1) accuracy (faces correctly labelled as CG or human), $d = 4.75$, $p = .005$, 95% CI [1.44, 8.06]; (2) discrimination (d'; of CG from human faces), $d = 1.93$, $p < .001$, 95% CI [0.99, 2.87]; and (3) dissimilarity ratings,[3] $d = 2.23$, $p < .001$, 95% CI [1.86, 2.59], were significantly above chance level. People were significantly more accurate at labelling human faces as human than labelling CG faces as CG (accuracy for human: $d = 4.97$, $p = .003$; accuracy for CG: $d = 4.50$, $p = .007$; $Q = 11.75$, $p < .001$). Both accuracy and d' were reduced for CG relative to human faces when stimuli were identity matched (accuracy: $d = 5.33$, $p = .004$; d': $d = 2.37$, $p < .001$) but not when they were unmatched (accuracy: $d = 0.26$, $p = .960$; d': $d = 1.31$, $p = .068$). However, moderation analyses indicated these differences between the identity matched and unmatched conditions was not significant (accuracy: $Q = 3.00$, $p = .083$; d': $Q = 0.08$, $p = .771$). Accuracy, but not d' or dissimilarity ratings, was significantly moderated by publication year (accuracy: $Q = 4.84$, $p = .028$, m[4] $= -1.58$; d': $Q = 0.22$, $p = .641$, $m = -0.18$; dissimilarity ratings: $Q = 0.30$, $p = .581$, $m = -0.25$), suggesting accuracy may be decreasing over time. However, neither accuracy, d' nor dissimilarity ratings were significantly moderated by realism difference (accuracy: $Q = 0.17$, $p = .684$, $m < 0.01$; d': $Q = 1.80$, $p = .180$, $m = 0.03$; dissimilarity ratings: $Q = 0.30$, $p = .585$, $m = 0.06$). Altogether, these results show people can CG faces apart from human ones (see Figs. 4 and 5).

Are CG faces perceived as less human-like?

As hypothesised (**H2**), meta-analysis found CG faces were perceived as less human-like than human faces, $d = 1.61$, $p < .001$, 95% CI [1.09, 2.13] (see Figs. 6 and 7). CG faces were perceived as significantly less human-like when stimuli were identity matched, $d = 1.97$, $p < .001$, but not when they were unmatched, $d = 0.82$, $p = .063$. However, moderation analyses indicated this difference between the identity matched and unmatched conditions was not significant, $Q = 2.24$, $p = .135$. Neither realism difference, $Q = 0.20$, $p = .657$, $m < 0.01$, nor publication year, $Q < 0.01$, $p = .976$, $m < 0.01$, significantly moderated effect sizes.

Is memory for CG faces poorer than for human faces?

As hypothesised (**H3**), meta-analysis found face memory was poorer for CG than human faces (see Figs. 8 and 9). This conclusion was supported by the results of three meta-analyses. First, recognition accuracy

[3] Studies using these ratings showed participants CG and human faces side-by-side and asked them to rate how similar the faces are (e.g., on a scale of 1 = not alike at all, 5 = totally alike: Gonzalez-Franco et al., 2016). Because a rating of 5 would indicate participants cannot tell CG and human faces apart, we used this value in the calculation of the one-sample effect size. For instance, for the familiar faces, pre-experiment condition in Gonzalez-Franco et al. (2016) the calculation was: $(3.12–5)/0.67 = -2.81$, which was then reversed to represent how dissimilar the faces were perceived to be.

[4] m refers to slope.

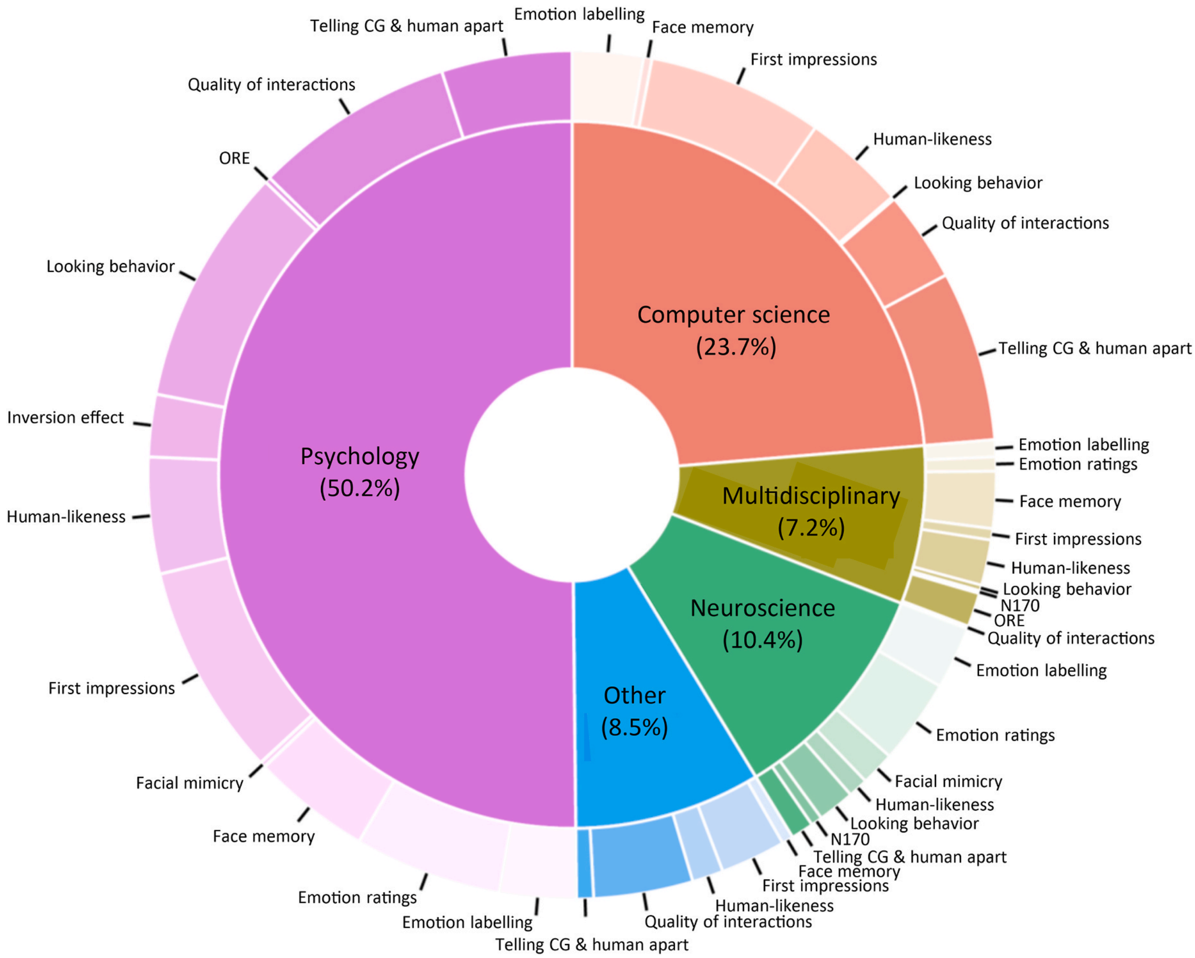

Fig. 3. Distribution of effect sizes across disciplines and face processing domains.
Note. Effect sizes were categorised into disciplines based on Journal Citation Reports (https://jcr.clarivate.com/).

(faces correctly judged as old or new) and sensitivity[5] (d': ability to tell new faces apart from previously learnt ones) was poorer for CG than human faces, $d = 0.58$, $p < .001$, 95% CI [0.37, 0.79]. Second, false alarms (i.e., misreports that a new, unfamiliar face has been seen before) were more common for CG than human faces, $d = 0.59$, $p < .001$, 95% CI [0.33, 0.85]. Poorer accuracy and increased false alarms persisted across both identity matched (accuracy: $d = 0.56$, $p < .001$; false alarms: $d = 0.52$, $p < .001$) and unmatched stimuli (accuracy: $d = 0.83$, $p = .003$; false alarms: $d = 0.74$, $p = .0006$) and moderation showed no difference between these levels (accuracy: $Q = 0.01$, $p = .930$; false alarms: $Q = 0.17$, $p = .680$). Neither publication year (accuracy: $Q = 0.78$, $p = .378$, $m = 0.02$; false alarms: $Q = 1.95$, $p = .163$, $m = -0.14$) nor realism difference (accuracy: $Q = 1.93$, $p = .165$, $m = 0.01$; false alarms: $Q = 0.29$, $p = .589$, $m < -0.01$) significantly moderated the size of effects. The third meta-analysis found no difference in reaction times (time taken to decide whether face is old or new) for CG and human faces, $d = 1.14$, $p = .215$, 95% CI [−0.66, 2.94], indicating that poorer memory and increased false alarms was not due to faster responding.

Is the quality of interactions with CG faces impoverished?

Contrary to our hypothesis (**H4**), meta-analysis found the quality of people's interactions with CG and human faces were not significantly different, though the effect was in the expected direction (i.e., worse for CG) and was borderline significant, $d = 0.33$, $p = .057$, 95% CI [−0.01, 0.67] (see Figs. 10 and 11). Moderation analysis showed realism difference significantly moderated the size of effects, $Q = 10.14$, $p = .001$, $m = -0.03$. This result is somewhat counterintuitive as it suggests the quality of interactions with CG and human faces are more alike when CG faces are *less* realistic relative to human ones. Neither identity matching, $Q = 0.01$, $p = .911$, nor publication year, $Q = 1.12$, $p = .289$, $m = -0.04$, significantly moderated effect sizes.

Do first impressions differ for CG and human faces?

As hypothesised (**H5**), meta-analysis found first impressions were significantly less favourable for CG than human faces, $d = 0.36$, $p = .005$, 95% CI [0.11, 0.60] (see Figs. 10 and 11). First impressions were significantly less favourable for CG than human faces when stimuli were matched for identity, $d = 0.51$, $p = .001$, but not when they were unmatched, $d = 0.09$, $p = .654$, However, moderation analyses indicated this difference between the identity matched and unmatched conditions was not significant, $Q = 1.94$, $p = .164$. Neither realism difference, $Q = 0.04$, $p = .850$, $m < 0.01$, nor publication year, $Q = 0.16$, $p = .691$, $m = -0.02$, moderated effect sizes. We also tested whether this effect replicated for trustworthiness ratings only, given their key role in first impressions research (Sutherland et al., 2013; Todorov et al., 2008). Contrary to overall first impressions, meta-analysis found there was no significant difference in trustworthiness ratings for CG and human faces, $d = 0.25$, $p = .251$, 95% CI [−0.18, 0.68].

[5] We kept d' separate from accuracy for telling CG and human faces apart because d' represented discrimination between CG and human faces. For face memory, both d' and accuracy represent discrimination between old (seen before) and new faces, so we analysed them together.

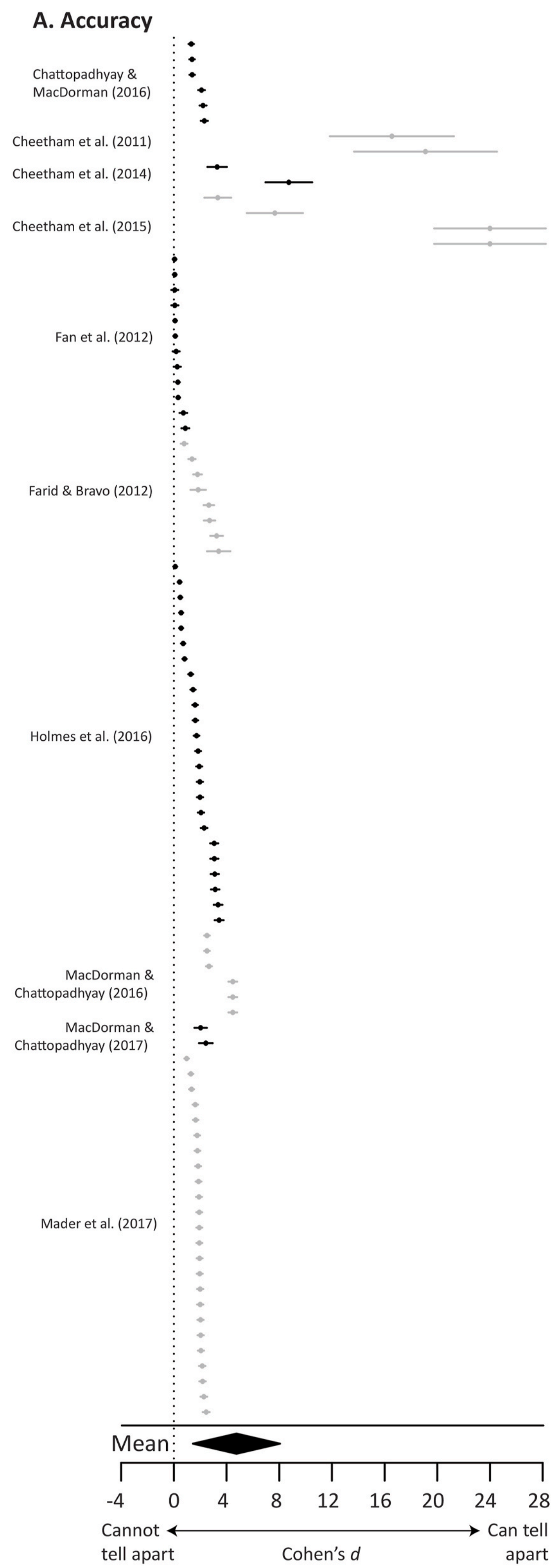

Fig. 4. Individual effect sizes for accuracy of telling CG and human faces apart.
Note. Effects may come from the same or different participants within each paper, or a mix of both. Error bars show 95% confidence intervals. We note that the outliers in studies by Cheetham and collegues in (A) are likely because they used stimuli which morphed between CG and human faces. We included only the endpoints of these continuua, where the standard deviation was quite small due to greater accuracy relative to more ambiguous stimuli.

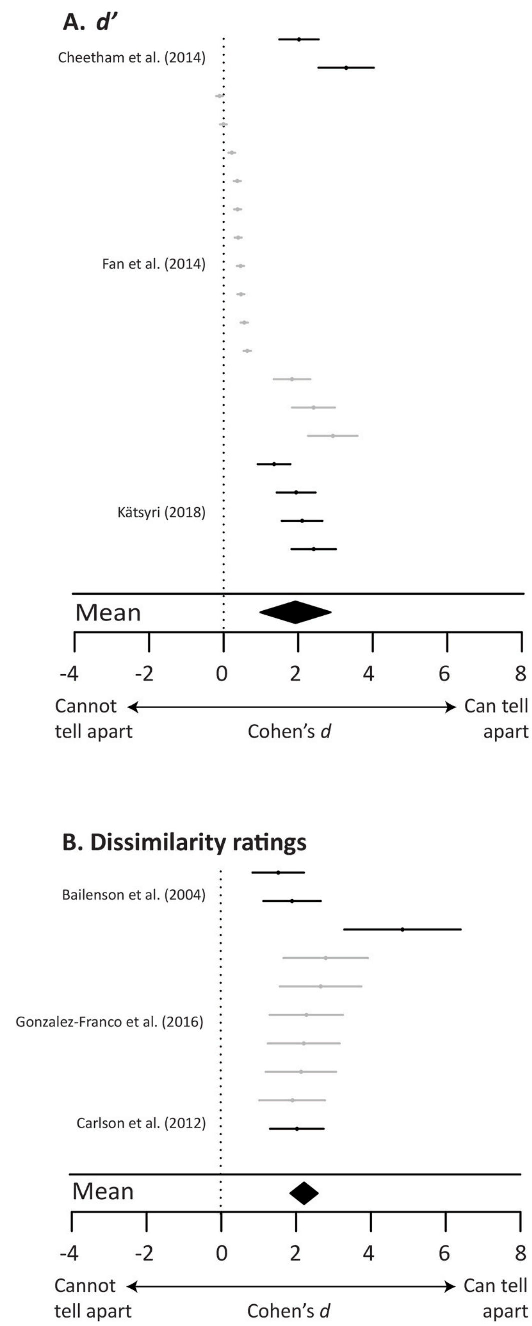

Fig. 5. Individual effect sizes for d' and similarity ratings for telling CG and human faces apart.
Note. Effects may come from the same or different participants within each paper, or a mix of both. Error bars show 95% confidence intervals.

Responses to emotional facial expressions

Meta-analysis found partial support for our hypothesis (H6) that emotion responses are reduced for CG relative to human faces (see Figs. 12 and 13). Emotion labelling was significantly less accurate for CG than human faces, $d = 0.34$, $p = .005$, 95% CI [0.11, 0.58], and while not significantly different, emotion ratings (i.e., intensity, arousal, valence), $d = 0.20$, $p = .150$, 95% CI [−0.07, 0.47], and facial mimicry (i.e., activation of congruent facial muscles), $d = 0.36$, $p = .281$, 95% CI [−0.29, 1.02], were reduced for CG compared to human faces. Emotion labelling was significantly less accurate for CG faces when: (1) stimuli were unmatched for identity, $d = 0.38$, $p = .005$ (vs matched: $d = 0.11$, $p = .744$), (2) expressions were moving, $d = 0.47$, $p = .012$ (vs non-

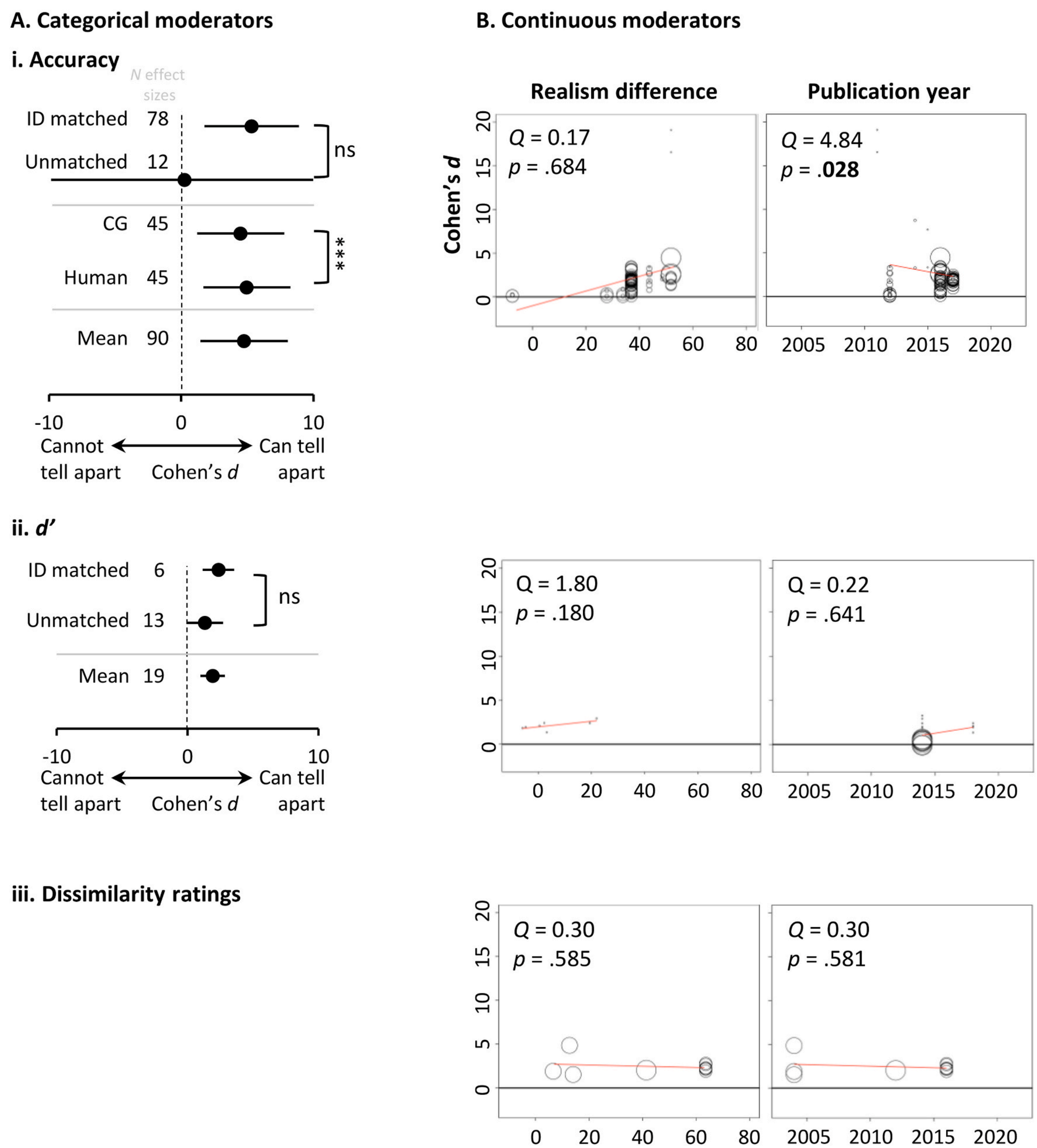

Fig. 6. Moderator analysis for telling CG and human faces apart.
Note. Error bars in (A) show 95% confidence intervals. ns indicates no significant difference between moderator levels. *** indicates $p < .001$. Dissimilarity ratings is missing a categorical moderator plot because all 10 effect sizes were identity matched.

moving: $d = 0.27$, $p = .051$), and (3) when displaying happiness, $d = 0.44$, $p = .030$, or when emotions were collapsed into a single effect size or showed an emotion other than a basic one (e.g., pain: Jackson et al., 2015), $d = 1.08$, $p = .0009$ (vs all other emotions $p > .101$). However, emotion labelling accuracy was not moderated by identity matching, $Q = 1.83$, $p = .176$, whether stimuli were moving or not, $Q = 1.59$, $p = .208$, or emotion category, $Q = 10.99$, $p = .139$. Emotion ratings, on the other hand, were significantly moderated by emotion category, $Q = 17.52$, $p = .008$, with CG faces eliciting reduced ratings for happy, $d = 0.55$, $p < .001$, and surprised expressions, $d = 0.72$, $p = .006$, relative to human faces (all other emotions $p > .344$). Facial mimicry for CG faces was weaker compared to human faces when stimuli were matched for identity, $d = 0.71$, $p < .001$ (vs unmatched: $d = 0.04$, $p = .570$), though this only included two papers. Neither publication year (both p's > 0.160) nor realism difference (both p's > 0.258) moderated emotion labelling or ratings. Overall, these findings indicate poorer responses to CG than human expressions.

Do people look at CG and human faces differently?

We combined frequency and duration of looking behaviour here because greater values for each indicate more time spent on that face. We had no specific hypothesis about looking behaviour, simply testing whether it differs for CG versus human faces. Meta-analysis revealed no difference in frequency and duration of looking at CG versus human faces, $d = 0.05$, $p = .545$, 95% CI [−0.11, 0.21] (see Fig. 14). Consistent with the overall meta-analysis, there was no difference for both identity matched, $d = 0.03$, $p = .825$, and unmatched stimuli, $d = 0.05$, $p = .682$, and moderation found no significant difference between these two levels, $Q = 0.10$, $p = .747$. There was also no difference in looking

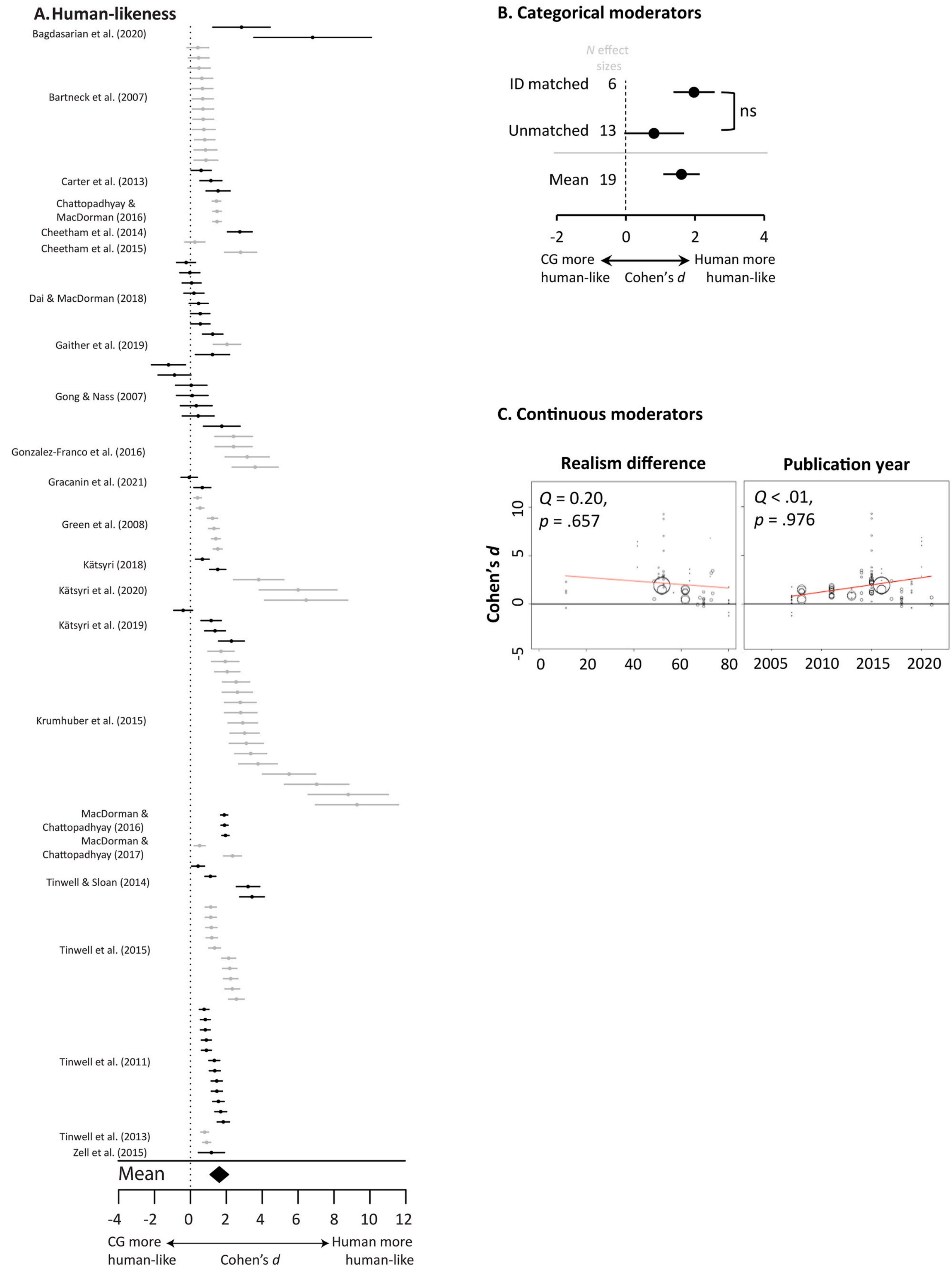

Fig. 7. Individual effect sizes and moderator analysis for human-likeness.
Note. Effects may come from the same or different participants within each paper, or a mix of both. Error bars in A. and B. show 95% confidence intervals. ns indicates no significant difference between moderator levels.

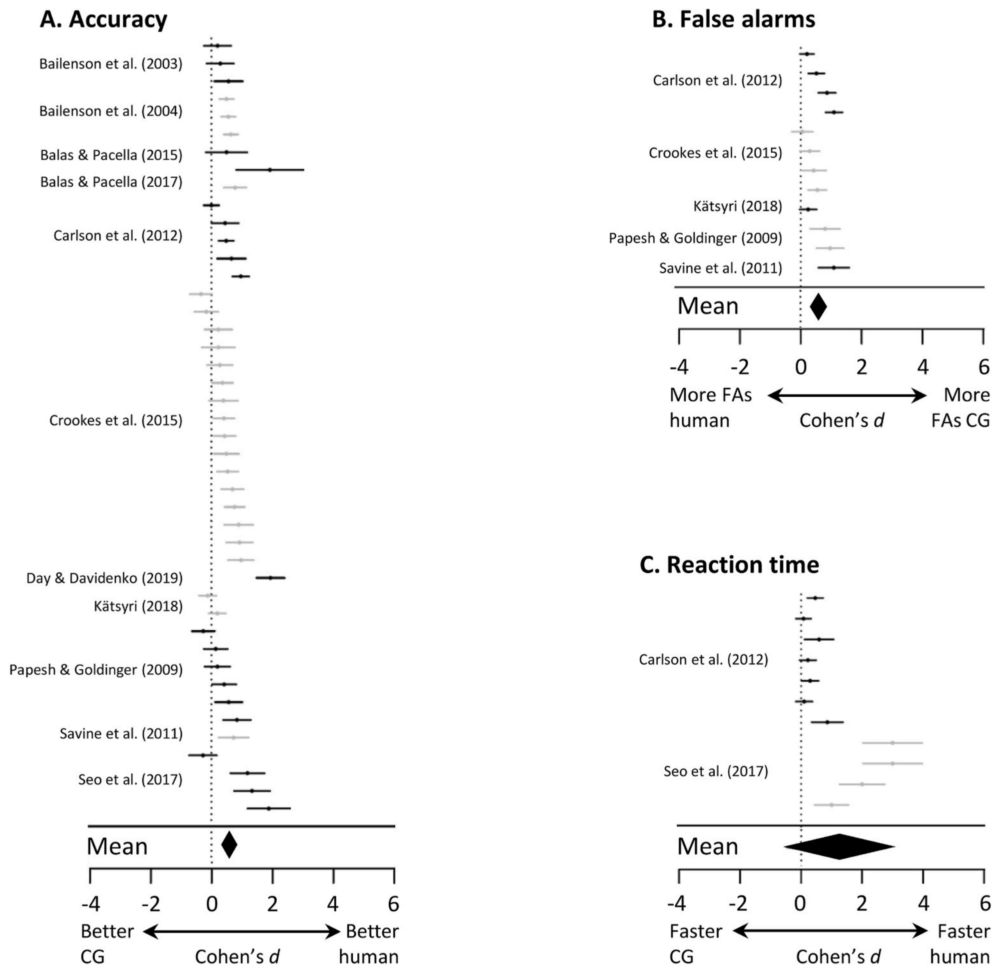

Fig. 8. Individual effect sizes for face memory.
Note. Effects may come from the same or different participants within each paper, or a mix of both. Error bars show 95% confidence intervals.

towards the CG versus human eyes,[6] $d = -0.17$, $p = .308$, and noses, $d < -0.01$, $p = .999$, though people did look significantly less at CG than human mouths, $d = 0.36$, $p = .017$. Realism difference, $Q = 4.90$, $p = .027$, $m = -0.04$, but not publication year, $Q = 0.06$, $p = .167$, $m = 0.06$, significantly moderated the size of effects, indicating a preference for looking at CG more than human faces when the realism difference between them was greater.

Hallmark face processing effects

Inversion effect. As hypothesised (H7), meta-analysis revealed a significant inversion effect for both CG, $d = 0.45$, $p = .009$, and human faces, $d = 0.80$, $p < .001$, with the size of the inversion effect significantly reduced for CG relative to human faces, $Q = 4.64$, $p = .031$ (see Fig. 15).

ORE. As hypothesised (**H8**), meta-analysis revealed a significant ORE for both CG, $d = 0.35$, $p = .011$, and human faces, $d = 0.45$, $p = .001$. However, the size of the ORE did not significantly differ CG relative to human faces, $Q = 0.74$, $p = .390$ (see Fig. 15).

N170. For the N170, we were only able to test whether the magnitude differed for CG versus human faces. Contrary to our hypothesis (**H9**), meta-analysis found no significant difference in the magnitude of the N170 for CG or human faces, $d = 0.12$, $p = .496$, 95% CI [−0.22, 0.46] (see Fig. 16). While this would indicate CG faces recruit face mechanisms to the same extent as human ones, we note that this analysis only included five effect sizes.

Narrative synthesis: EEG/ERP

Our review also uncovered eleven papers comparing EEG/ERP responses other than the N170 for CG and human faces. Across these studies, nine comparisons found the target EEG response was decreased for CG compared to human faces (P100: Mühlberger et al., 2009, P200: Gonzalez-Franco et al., 2016; Mustafa & Magnor, 2016; N400: Mustafa et al., 2017; LPP: Cheetham et al., 2015; Sollfrank et al., 2021; harmonic responses to an FPVS paradigm, for medial occipital ROI: Gwinn et al., 2018; alpha & theta dB for fear expressions, Sollfrank et al., 2021), four found the target EEG response was decreased for CG compared to human faces (N300: Sollfrank et al., 2021; LPP for people with low social anxiety: Mühlberger et al., 2009; Schindler et al., 2017; harmonic responses to an FPVS paradigm, right occipito-temporal ROI: Gwinn et al., 2018), and eight found the target EEG response was not significantly different for CG and human faces (N100: Sollfrank et al., 2021, P300: Seo et al., 2017; N400 for male stimuli: Mustafa et al., 2017; EPN: Mühlberger et al., 2009; Schindler et al., 2017; LPP for people with high social anxiety: Mühlberger et al., 2009; alpha & theta dB for neutral faces: Sollfrank et al., 2021). Overall, these results demonstrate that EEG responses to CG relative to human faces are mixed.

[6] We could not include face region in the full model because it meant excluding all 64 effect sizes from Bucher & Voss (2019) which altered results for other moderators.

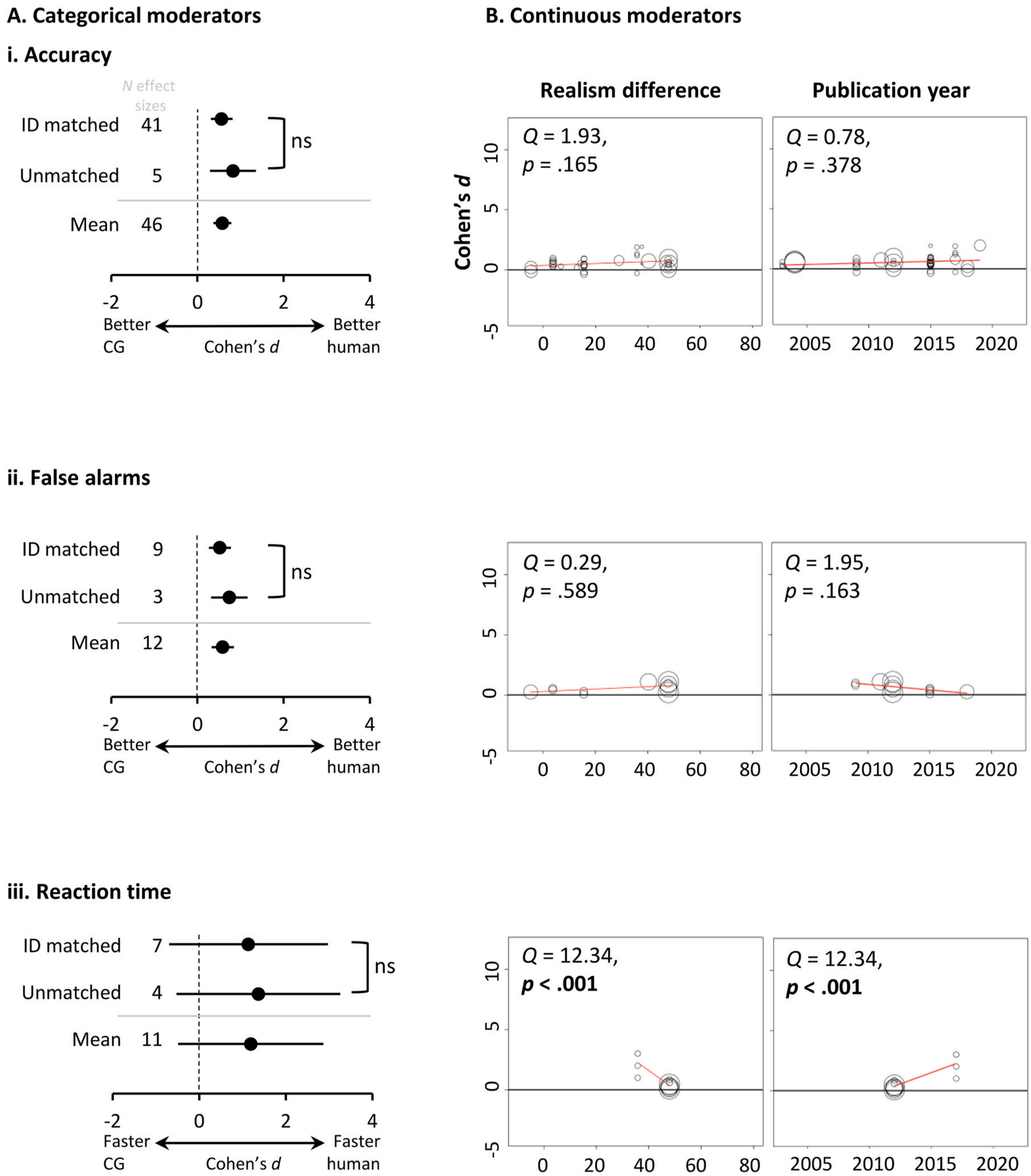

Fig. 9. Moderator analysis face memory.
Note. Error bars in (A) show 95% confidence intervals. ns indicates no significant difference between moderator levels.

Narrative synthesis: brain imaging

Meta-analysis could not be conducted for brain imaging because, while most studies provided data for x, y co-ordinates, data for BOLD responses was only available for two studies (Freeman et al., 2014; Kätsyri et al., 2020). The FFA is a brain region that responds more strongly to faces than other objects (Kanwisher et al., 1997). Despite the importance of the FFA to face processing (i.e., known to respond more strongly to faces than other objects: Kanwisher et al., 1997), only two studies have compared FFA activation for CG and human faces, both reporting reduced FFA activation for CG relative to human faces (Kätsyri et al., 2020; Mundy et al., 2012). Results for the range of other brain regions tested are inconsistent. Three studies reported reduced activation for CG compared to human faces, for the superior temporal sulcus and left inferior frontal gyrus (fear expressions: Kegel et al., 2020), fusiform gyri, cerebellum, left superior temporal gyrus & rectal gyri (Moser et al., 2007), and dorsomedial prefrontal cortex, rostral anterior cingulate cortex, ventromedial prefrontal cortex (Riedl et al., 2014). Two studies reported increased activation for CG relative to human faces, for the left medial temporal gyrus (Moser et al., 2007) and the perirhinal cortex (Mundy et al., 2012). Nine comparisons indicated no difference in response activation to CG versus human faces, particularly for the amygdala (Freeman et al., 2014; Kätsyri et al., 2020; Kegel et al., 2020; Moser et al., 2007), but also for the superior temporal sulcus and left inferior frontal gyrus (neutral expressions: Kegel et al., 2020), lateral occipital complex, parahippocampal place area and posterior hippocampus (Mundy et al., 2012). Two additional studies using fNIRS also found mixed results, with one reporting reduced activation in the

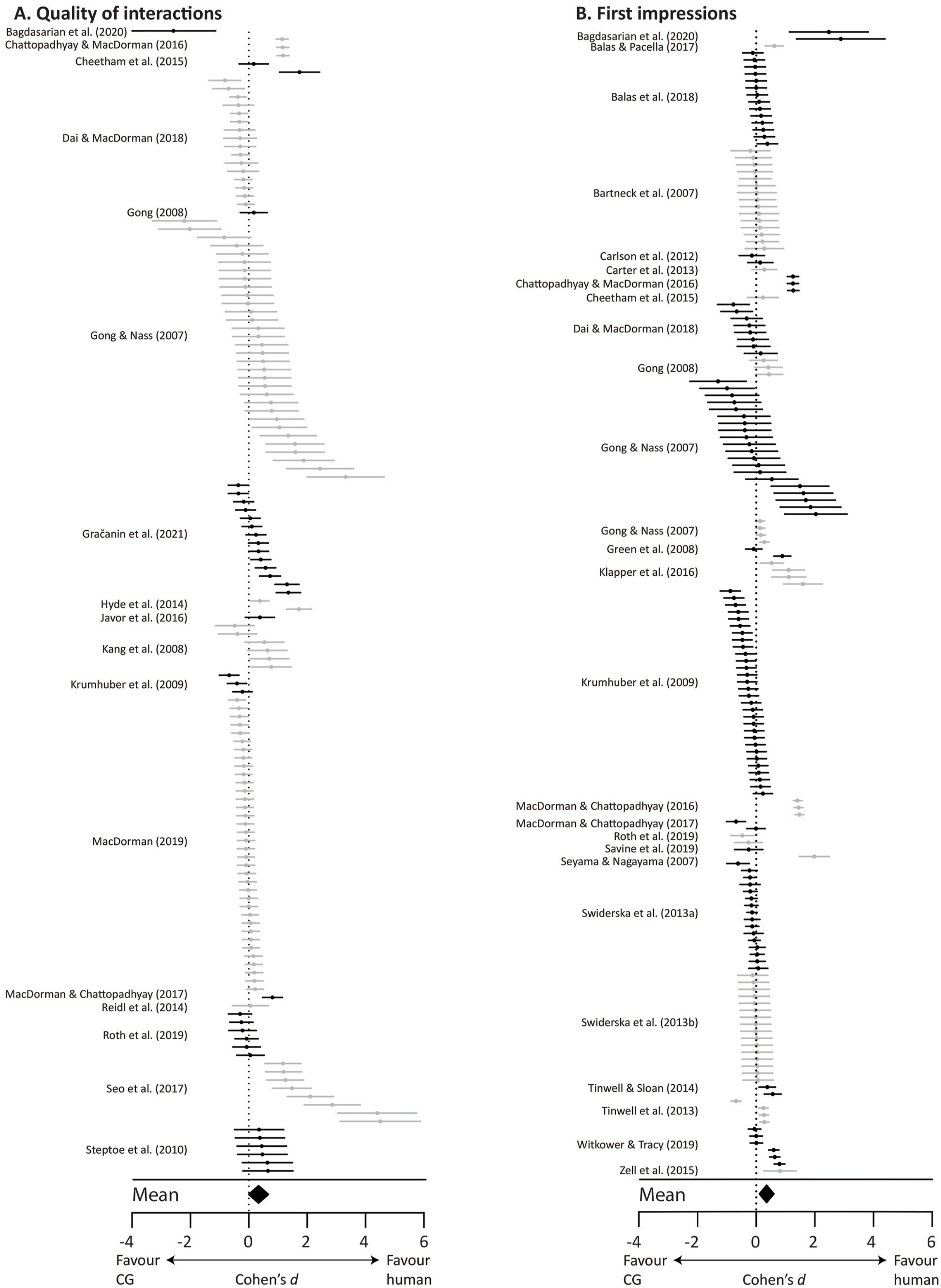

Fig. 10. Individual effect sizes for quality of interactions and first impressions.
Note. Effects may come from the same or different participants within each paper, or a mix of both. Error bars show 95% confidence intervals.

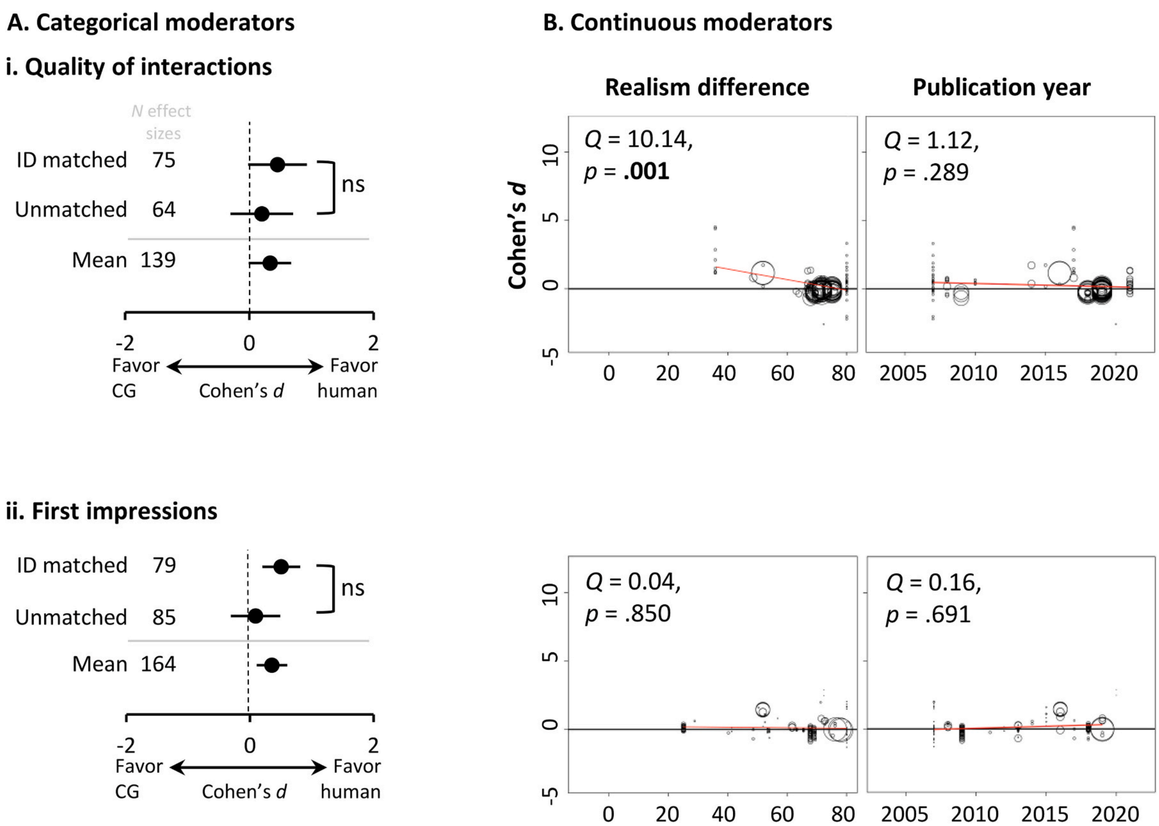

Fig. 11. Moderator analysis for quality of interactions and first impressions.
Note. Error bars in (A) show 95% confidence intervals. ns indicates no significant difference between moderator levels.

orbitofrontal cortex (Zhao et al., 2020) and the other finding no difference in the dorsolateral prefrontal cortex (Wang, Xu, et al., 2019) for CG compared to human faces.

Narrative synthesis: perceptions of face gender, sexuality, and race

There were a small number of studies investigating people's judgements of face gender (3 studies), race (2 studies), and sexuality (1 study). Large methodological differences between these studies made meta-analysis inappropriate. Overall, it is unclear whether gender, sexuality, and/or race judgements differ for CG and human faces. Balas and Pacella (2015) found gender categorisation was less accurate but faster for CG than human faces, and Gaither et al. (2019) reported that the social phenomenon of hypodescent, whereby people who appear racially ambiguous are automatically categorised into the minority race, is reduced for CG relative to human faces. However, two studies found people's mouse movement patterns were similar for CG and human faces when categorising face gender (Freeman et al., 2008) and race (Freeman et al., 2010b), and two other studies investigated relevant effects, but did not compare performance for CG and human faces. Freeman et al. (2010a) found masculinised female faces and feminised male faces were more likely to be judged as homosexual for both CG and human faces, but did not compare the size of these effects, and Green et al. (2008) investigated the extent to which faces are perceived as feminine, but also did not compare performance for CG and human faces.

Narrative synthesis: speech identification and learning

Three studies have investigated speech in CG compared to human faces. However, large methodological differences between these studies made meta-analysis inappropriate. Of two studies investigating speech identification, one (Lidestam et al., 2001) found ability to identify speech was worse for CG than human faces, and the other (Gibert et al., 2013) did not compare performance for CG and human faces. A third study (Chen et al., 2019) found people learned to pronounce speech better when instructed by CG than human faces in some conditions (consonants and atypical vowels) but not others (typical vowels).

Publication bias analysis

We tested for evidence of publication bias for each meta-analysis, using the residuals of the final models to account for dependency amongst effect sizes (e.g., multiple effects from the same participants). Egger's regression test (Egger et al., 1997) found no significant asymmetry of the residuals for any meta-analysis (all p's > 0.09), indicating there was no evidence of publication bias. See Supplement S5 for funnel plots and full publication bias analysis.

Discussion

The present review highlights the multidisciplinary nature of work on CG faces, showing large contributions from psychology, neuroscience, and computer science (see Fig. 3). Our synthesis of this multidisciplinary evidence established that, for the types of stimuli used to date, people are clearly able to tell CG and human faces apart and perceive CG faces as less human-like than human faces. Several other types of responses also favoured human over CG faces, including memory for faces, first impressions, and labelling of emotional expressions, and there was evidence hallmark face effects such as the inversion effect, although present, are weaker for CG than for human faces. However, there were also some circumstances where people's responses to CG and human faces do not seem to differ, including quality of interactions, ratings of emotional expressions, facial mimicry, and looking behaviour. However, the lack of significant differences in some instances may be due to a

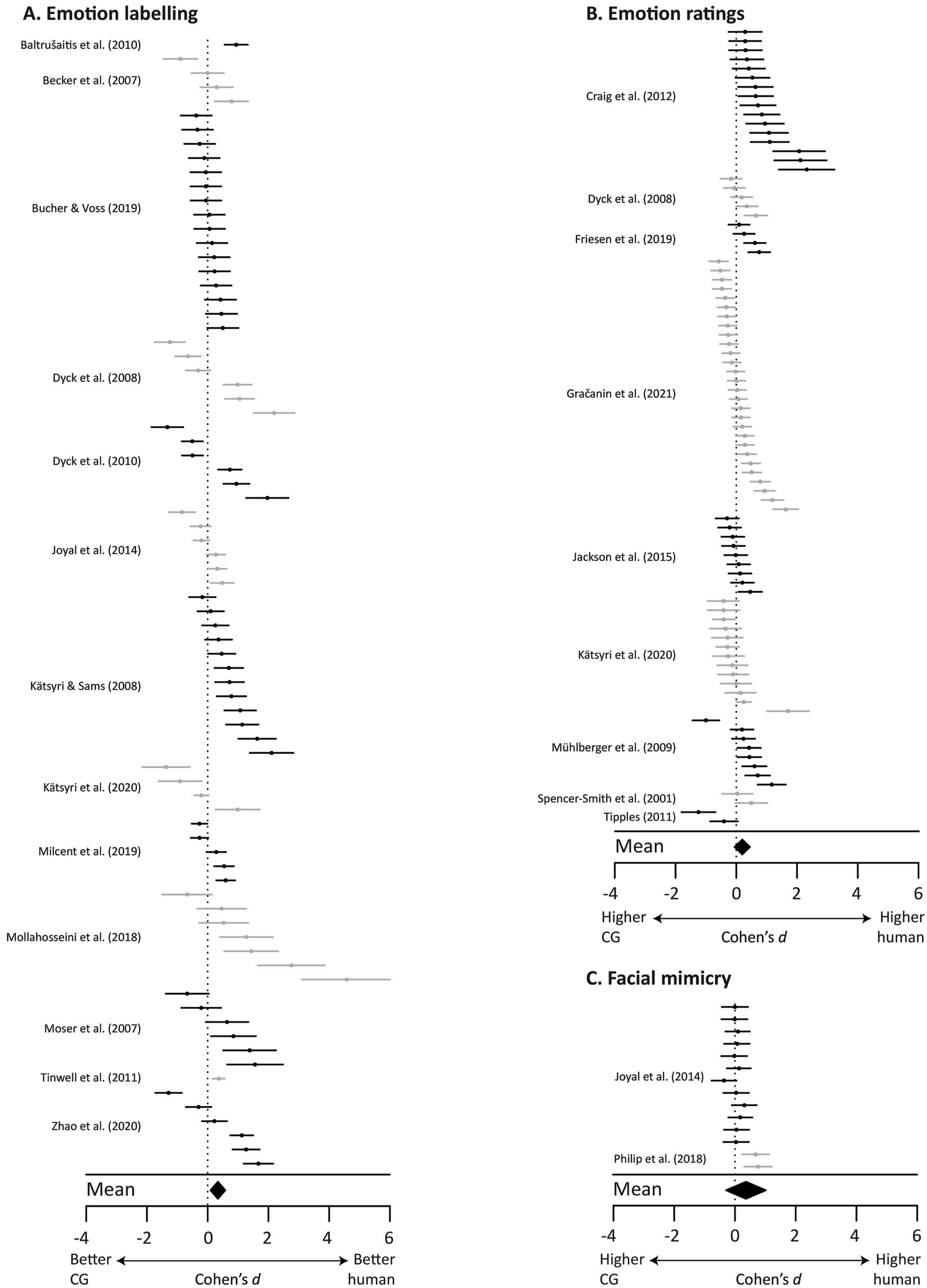

Fig. 12. Individual effect sizes for responses to facial expressions.
Note. Effects may come from the same or different participants within each paper, or a mix of both. Error bars show 95% confidence intervals.

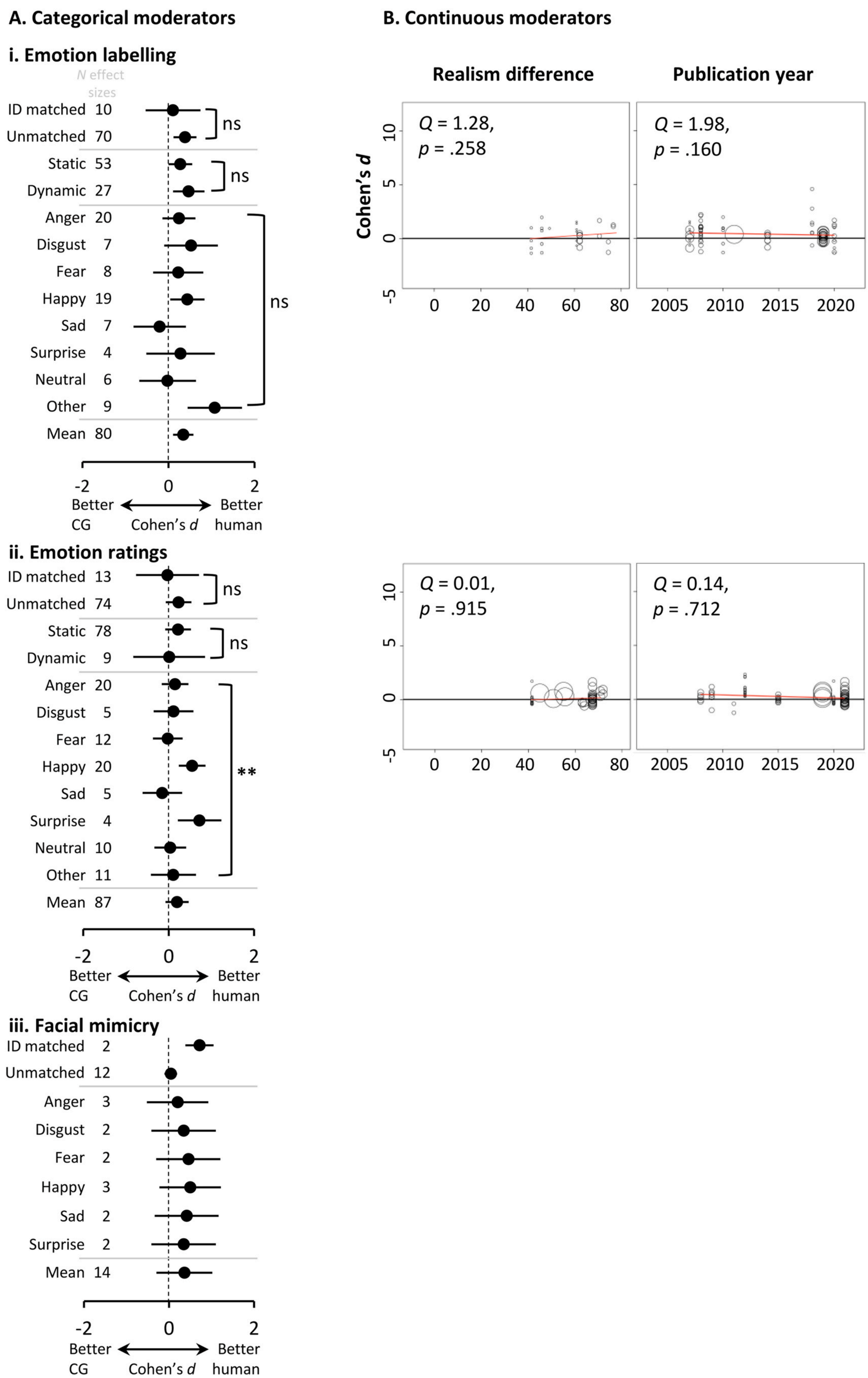

Fig. 13. Moderator analysis for responses to facial expressions.

Note. Error bars in (A) show 95% confidence intervals. ns indicates no significant difference between moderator levels. ** indicates $p < .01$. We were unable to run moderators for facial mimicry because the analysis only included two papers.

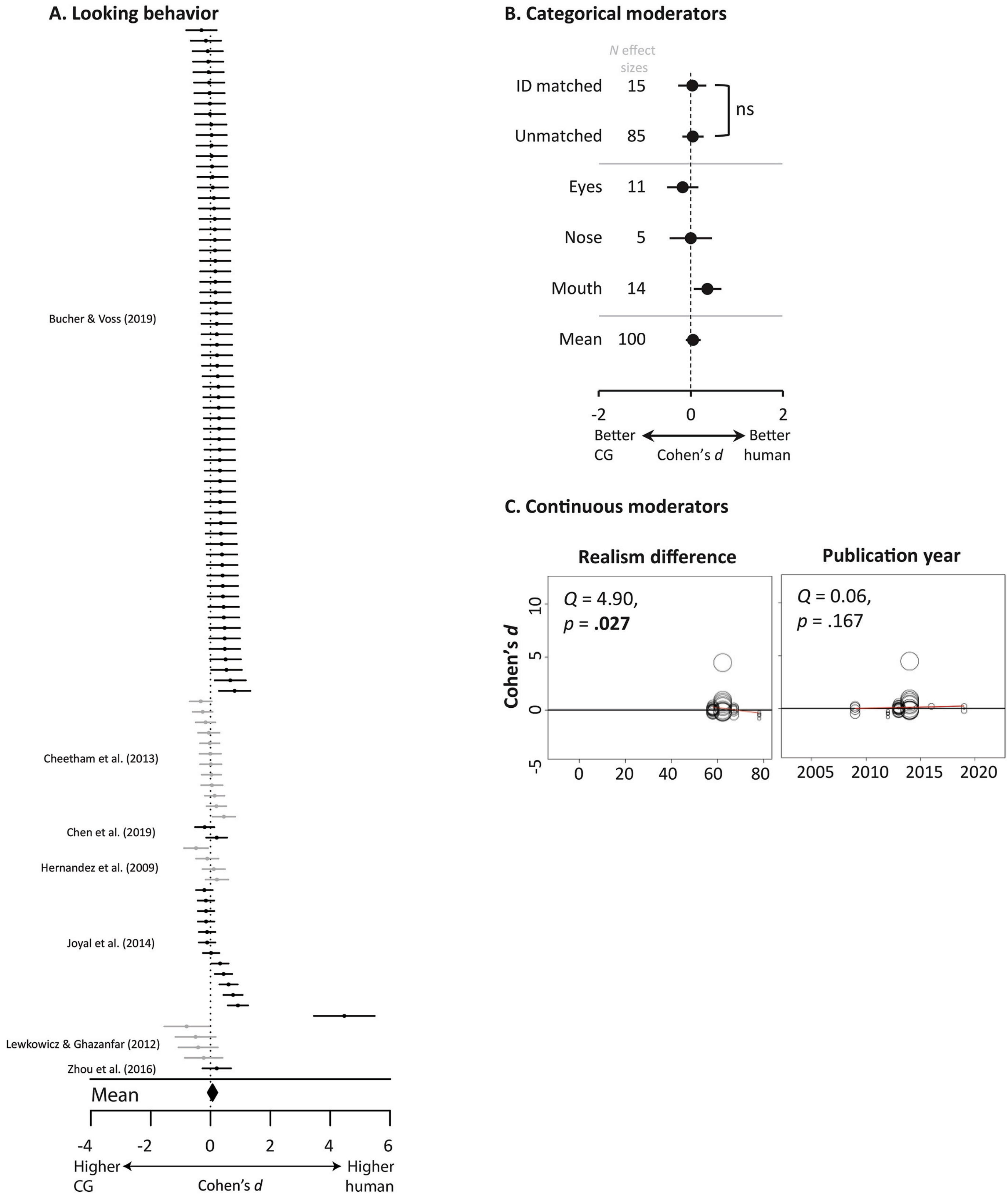

Fig. 14. Individual effect sizes and moderator analysis for looking behaviour.
Note. Effects may come from the same or different participants within each paper, or a mix of both. Error bars in A. and B. show 95% confidence intervals. ns indicates no significant difference between moderator levels. We were unable to run moderator analysis for face region (eyes, nose, mouth).

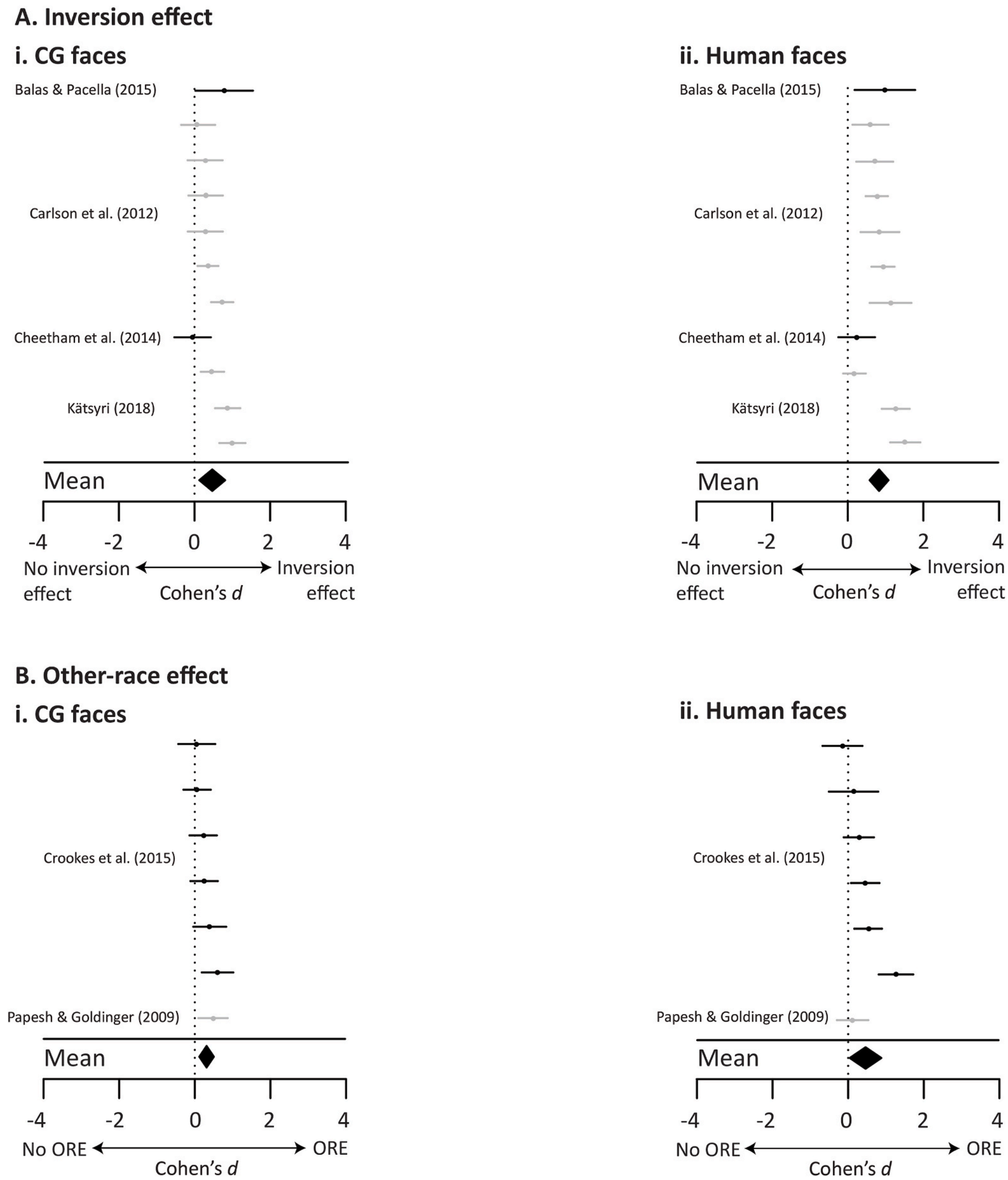

Fig. 15. Individual effect sizes for the inversion and other-race effects.
Note. Effects may come from the same or different participants within each paper, or a mix of both. Error bars show 95% confidence intervals.

lack of data within a domain. For example, small-medium effects for ratings of emotional expressions and facial mimicry may indicate these responses are also impoverished, but further data are needed to establish effect reliability. We also found no evidence of publication bias in any face domain, indicating a lack of bias in this growing research area. Altogether, these findings provide deeper insight into when people's responses to CG faces are most likely to be impoverished, which can be used to inform how we apply and engage with CG beings across health, social, and scientific settings.

Telling CG and human faces apart and perceptions of human-likeness

Our findings that people are clearly able to tell CG and human faces apart and perceive CG faces as less human-like than human faces, as indicated by large meta-analytic effect sizes, are consistent with the avatar-feature hypothesis (Cheetham et al., 2013). The avatar-feature hypothesis suggests that people are highly sensitive to features of non-humanness, such as lack of animacy in CG skin and eyes (Balas & Horski, 2012; Balas & Tonsager, 2014; MacDorman et al., 2009; Vaitonytė et al., 2021) and lack of fine-grained texture in the skin (Itz et al., 2014). However, our findings contrast with emergent evidence that people are unable to tell the latest hyper-realistic AI-generated faces apart from real human faces (Nightingale & Farid, 2022; Shen et al., 2021). Even with extensive training, people's ability to tell these AI-generated faces apart from human ones only reaches levels just above chance (Nightingale & Farid, 2022). However, just because most CG faces used to date are clearly identifiable as non-human does not mean they are void of value in real-life applied settings. There are some instances where people may benefit from interacting with a clearly CG being, such as when anonymity is preferred (e.g., Shepherd & Edelmann, 2005). Just as people disclose more information to a "stranger on a train/plane" when meeting them again is unlikely (John et al., 2011), CG faces low in realism may increase feelings of anonymity by assuming the face is controlled by a computer rather than a human, leading to a better interaction (Pickard et al., 2016). Our finding that interaction quality is not necessarily impoverished for CG compared to human faces supports this idea. Lower-realism CG faces also bypass ethical concerns

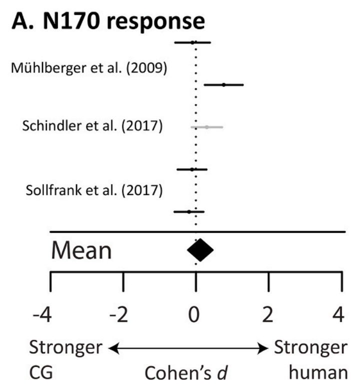

Fig. 16. Individual effect sizes for the N170 response.
Note. Effects may come from the same or different participants within each paper, or a mix of both. Error bars show 95% confidence intervals.

presented by virtual beings that fool people into believing they are real humans. For instance, concerns have been raised about potential for misuse in scams and other types of deception, including election fraud (de Ruiter, 2021) and non-consensual creation of pornographic material using one's face (Maras & Alexandrou, 2019), where people are not aware that a hyper-realistic CG face has been used instead of a human one.

In a similar vein, regardless of the accessibility of hyper-realistic AI-generated faces for empirical research (e.g., one can purchase 1000 faces for $1 at https://generated.photos), lower-realism CG faces remain a useful tool for experiments investigating face perception. Lower-realism CG faces afford excellent experimental control because specific parameters of interest can be fine-tuned in ways that are not yet possible for GAN or human faces (e.g., adding or removing subtle facial actions from a single identity face [Malek et al., 2019]) or systematically manipulating faces on traits like trustworthiness [Todorov et al., 2008]). Thus lower-realism CG faces can have an important role in testing causative relationships, which requires high experimental control, and seeking convergent evidence for hypotheses originating from studies of naturalistic faces.

Our finding that interaction quality is not necessarily impoverished for CG compared to human faces also supports the inherent value of low-realism CG faces. Given people's empathic responses to other clearly non-human agents (e.g., pets: Gómez-Leal et al., 2021), it is perhaps unsurprising that we can empathise with faces designed to appear human-like. We also found that interactions with CG faces were rated more similar to human ones as CG faces became less realistic. It is possible that less realistic CG faces are more predictable and therefore elicit a more enjoyable interaction, whereas more realistic interaction partners err towards the uncanny valley (Mori, 1970) and elicit feelings of unease. Alternatively, it is possible that less realistic CG faces are perceived as more anonymous because they appear less like a human who can judge. Due to these ethical concerns and potential benefits, some platforms are choosing to engage CG beings that are clearly non-human, including the World Health Organisation's digital health workers (https://www.paho.org/en/alcohol/pahola; https://www.soulmachines.com/2020/09/florence-digital-health-worker/).

Impoverished responses for CG faces

Meta-analyses found the apparent differences between CG and human faces also played out in impoverished responses to CG stimuli across several key aspects of face processing, including memory, the face inversion effect, first impressions, and emotion labelling. Memory was poorer for CG than human faces, as evidenced by medium effect sizes for accuracy and false alarms, coupled with no significant difference in reaction times (i.e., no evidence of speed-accuracy trade-off). This evidence of poorer memory for CG faces may indicate that individual CG identities are less differentiated in face space than human ones (Valentine, 1991; Valentine et al., 2016), potentially owing to lack of features used to discriminate human faces such as facial texture and reflectance (Itz et al., 2014), or lack of childhood exposure to CG faces (McKone et al., 2019; Singh et al., 2021). Our finding of poorer memory for CG than human faces also dovetails with evidence that the inversion effect, an established hallmark of human face processing, is reduced for CG faces. While initial evidence suggests that greater exposure to CG faces does not increase CG face expertise (Crookes et al., 2015), this exposure is likely to have occurred after the critical period for development of face expertise (McKone et al., 2019; Singh et al., 2021). This leaves open the question of whether CG faces will increasingly engage human face processing mechanisms over generations, as people's childhood "face diet" expands to include CG faces (e.g., via virtual influencers like Lil Miquela).

Meta-analyses also found first impressions were less favourable for CG than human faces, as evidenced by small to medium effects. However, it is unclear precisely which types of first impressions are impoverished. There were enough trustworthiness rating studies available to analyse separately, revealing no significant CG-human difference in this aspect of first impressions. It may be that CG faces are perceived as trustworthy for different reasons than human ones, potentially contributing to the lack of CG-human differences we found for interaction quality. Indeed, the recent work on AI-generated faces indicates they may be perceived as more trustworthy than human ones (Nightingale & Farid, 2022). We were unable to produce a more fine-grained analysis of other rating types due to the large variety. Future research would benefit from systematically testing different types of first impressions, to develop a theoretical framework that can predict what types of responses will be impoverished and why.

Finally, meta-analysis found a small-medium effect showing emotion labelling was less accurate for CG than human faces. There are several potential reasons for this finding. CG faces lack physical cues usually present in human expressions which aid emotion identification, such as colour (Thorstenson et al., 2019), tears (Gračanin et al., 2021), and pupil changes (Milcent et al., 2019). The absence of these peripheral emotion cues may increase reliance on individual muscle movements (i.e., action units; Ekman, 1992) that may be poorly or unrealistically replicated in CG expressions. For instance, the Duchene marker (Miller et al., 2020) and nose wrinklier (Dyck et al., 2008) in happy and disgust expressions, respectively, are difficult to reproduce on CG faces, and open-mouthed expressions with teeth (e.g., smiles) often appear strange. For dynamic expressions, researchers may inadvertently combine action units or use timings which do not occur in real life, making them more difficult to identify (Dawel et al., 2021). In contrast to these differences for emotion labelling, we found no significant CG-human difference for ratings of emotion related phenomena (valence, intensity, arousal) or facial mimicry responses. The finding of equivalent mimicry may reflect people's tendency to anthropomorphise may be facilitating empathy for CG faces as for human ones (Roth et al., 2019). However, it is unclear how robust this conclusion is, as we still found small-medium effects pointing to impoverished responses for CG faces.

Implications for face science

We found that CG faces elicited face specific N170 responses, and inversion (Yin, 1969) and other-race (Meissner & Brigham, 2001) effects. While the magnitude of the ORE and N170 responses (Bentin et al., 1996; Rossion & Jacques, 2012) were the same for CG and human faces, the inversion effect was reduced for CG relative to human faces. However, the lack of data comparing these hallmark effects for CG versus human faces is concerning. There has been a clear increase in the number of studies using CG faces as stimuli to investigate human face processing (Dawel et al., 2021), and the present review highlights that this has occurred in absence of a thorough understanding of whether CG faces replicate important face-specific effects relative to human faces. While many studies have tested hallmark face effects in CG faces alone (e.g., Geiger & Balas, 2021; Hourihan et al., 2013; Matheson & McMullen, 2011), more direct comparison to human ones are required to determine the extent to which CG faces engage face-specific processing mechanisms.

An additional domain of interest in face science is the characteristic ways in which people look at or scan faces. We found that there was no difference in looking frequency and duration for CG versus human faces. However, our review leaves open the question of whether scanning patterns differ for CG and human faces (also see Supplement S6 for narrative review of looking behaviour studies). The eye region is central to face recognition (Royer et al., 2018; Schyns et al., 2003) and emotion processing (Calvo et al., 2018) in human faces and animacy determination in CG faces (Balas & Horski, 2012; MacDorman et al., 2009; Vaitonytė et al., 2021). However, it is possible that the pattern of exploration for CG faces differs from human ones (e.g., people may look to combine features in different ways). If hyper-realistic CG faces (Karras et al., 2021) continue to infiltrate our social world as we predict they will, eye-tracking technology via webcams could potentially identify where scanning patterns differ and alert the user to non-humanness in CG faces, where they appear consciously indistinguishable from human ones (Nightingale & Farid, 2022; Shen et al., 2021). It is therefore important for future research to investigate how CG and human faces are explored in terms of overt looking behaviour, particularly for the newer, AI-generated CG faces (Karras et al., 2021).

Finally, several meta-analyses revealed that effects were larger for identity matched than non-matched stimuli. This result suggests that not matching CG and human face identities may inadvertently hide differences between them, highlighting the importance of identity matching in scientific research.

Implications for real life

The findings of our meta-analyses have important practical implications for the use of CG faces in real-life and suggest that CG faces may be appropriate substitutes in some situations but not others. For instance, virtual influencers like Lil Miquela may be just as effective at brand advertising as human influencers, given our results would suggest we trust and interact with her as we would for humans. However, other scenarios, like e-therapy, may be inappropriate places for CG faces, as having less favourable first impressions and being unable to identify emotional expressions may have severe negative consequences, because it suggests that CG faces are unlikely to bolster our wellbeing to the same extent as human-to-human relationships (Cruwys et al., 2014). As such, CG faces should be used to supplement rather than replace human ones (e.g., accessing a CG therapist for non-urgent after-hours support, when human ones are unavailable). Our results concerning identity matching also suggest that CG therapists should have their own individual identity. Responses to a CG therapist might be poorer if it is made to appear like the real human therapist, because being able to make a direct comparison with a human-to-human experience exaggerates that the interaction with their CG version is impoverished.

Limitations

The present results are founded on a range of CG face realism with few AI-generated studies included, as AI-generated face studies are only just beginning to emerge in the scientific literature (Dawel et al., 2021). Single study results for the latest hyper-realistic AI-generated faces indicate they may be indistinguishable from human faces (Nightingale & Farid, 2022; Shen et al., 2021). The present review highlights the breadth of domains that AI-generated faces will need to be tested to fully understand human responses to them. The scope of the present review is also constrained to studies directly comparing responses to CG and human faces. It may be useful for future reviews to synthesise studies that have tested only CG faces. For instance, some studies have tested hallmark face processing effects using just CG faces (Geiger & Balas, 2021; Hourihan et al., 2013; Kala et al., 2021; Matheson & McMullen, 2011). This body of work could be used to establish the robustness of these effects in CG faces. Finally, the present review focused on faces because of their central role in social interaction. However, CG faces, like human ones, are often presented with other context, such as verbal or written content, or bodies. Contextual factors have a strong influence on responses to human faces (Aviezer et al., 2008; Deffler et al., 2015; Gendron et al., 2013; Wieser & Brosch, 2012) and it is reasonable to expect they would also influence responses to CG faces.

Conclusion

The results of the present systematic review and meta-analyses reveal several ways in which responses to CG and human faces differ. While responses to hyper-realistic AI-generated faces may be enhanced in some instances (e.g., higher trustworthiness than for human faces; Nightingale & Farid, 2022), we found that no type of response was better or enhanced for CG relative to human faces overall. Few studies have sought to compare hallmark face processing for CG and human faces, and therefore caution should be applied when using CG faces to study these effects. For real-life applications, a common assumption is that more realistic CG faces will produce better outcomes. This is not necessarily the case (Parmar et al., 2022). Low realism CG beings are effective at providing support and increasing wellbeing in a range of settings, such as for reducing alcohol use (Rubin et al., 2022, see also: https://www.paho.org/en/alcohol/pahola), reducing shame (Christensen et al., 2013), and improving bedside manner of doctors (Andrade et al., 2010, see also: https://medicalcyberworlds.com). CG beings that look clearly CG may also be preferred because of benefits associated with online anonymity for some groups (e.g., socially anxious people; Shepherd & Edelmann, 2005) and to avoid the ethical implications of deep fakes (de Ruiter, 2021).

Author note

This research is supported by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (project DP220101026) and by a TRANSFORM Career Development Fellowship to AD from The Australian National University (ANU) College of Health and Medicine. The funders had no role in developing or conducting this research. We have no conflicts of interest to disclose.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Full data and code are available here: https://osf.io/fxcku/.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbr.2023.100283.

References

Abbott, M. B., & Shaw, P. (2016). Virtual nursing avatars: Nurse roles and evolving concepts of care. *Online Journal of Issues in Nursing, 21*(3), 1–6. https://doi.org/10.3912/OJIN.Vol21No03PPT39,05

Afzal, S., Sezgin, T. M., Gao, Y., & Robinson, P. (2009). Perception of emotional expressions in different representations using facial feature points. *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 1–6. https://doi.org/10.1109/ACII.2009.5349549

Andrade, A. D., Bagri, A., Zaw, K., Roos, B. A., & Ruiz, J. G. (2010). Avatar-mediated training in the delivery of bad news in a virtual world. *Journal of Palliative Medicine, 13*(12), 1415–1419. https://doi.org/10.1089/jpm.2010.0108

Arsalidou, M., Morris, D., & Taylor, M. J. (2011). Converging evidence for the advantage of dynamic facial expressions. *Brain Topography, 24*(2), 149–163. https://doi.org/10.1007/s10548-011-0171-4

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., & Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science, 19*(7), 724–732. https://doi.org/10.1111/j.1467-9280.2008.02148.x

Bailenson, J. N., Beall, A. C., & Blascovich, J. (2003). Using virtual heads for person identification: An empirical study comparing photographs to photogrammetrically-generated models. *Journal of Forensic Identification, 53*(6), 722–728.

Bailenson, J. N., Beall, A. C., Blascovich, J., & Rex, C. (2004). Examining virtual busts: Are photogrammetrically generated head models effective for person identification? *Presence: Teleoperators and Virtual Environments, 13*(4), 416–427. https://doi.org/10.1162/1054746041944858

Balas, B., & Horski, J. (2012). You can take the eyes out of the doll, but. *Perception, 41*(3), 361–364. https://doi.org/10.1068/p7166

Balas, B., & Pacella, J. (2015). Artificial faces are harder to remember. *Computers in Human Behavior, 52*, 331–337. https://doi.org/10.1016/j.chb.2015.06.018

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., & Tonsager, C. (2014). Face animacy is not all in the eyes: Evidence from contrast chimeras. *Perception, 43*(5), 355–367. https://doi.org/10.1068/p7696

Balas, B., Westerlund, A., Hung, K., & Nelson Iii, C. A. (2011). Shape, color and the other-race effect in the infant brain. *Developmental Science, 14*(4), 892–900. https://doi.org/10.1111/j.1467-7687.2011.01039.x

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America - PNAS, 104*(46), 17948–17953. https://doi.org/10.1073/pnas.0705435104

Balsters, M. J. H., Krahmer, E. J., Swerts, M. G. J., & Vingerhoets, A. J. J. M. (2013). Emotional tears facilitate the recognition of sadness and the perceived need for social support. *Evolutionary Psychology, 11*(1), 148–158. https://doi.org/10.1177/147470491301100114

Baltrušaitis, T., Riek, L. D., & Robinson, P. (2010). Synthesizing expressions using facial feature point tracking: How emotion is conveyed. *Proceedings of the 3rd International Workshop on Affective Interaction in Natural Environments*, 27–31. https://doi.org/10.1145/1877826.1877835

Bartneck, C. (2001). How convincing is Mr. Data's smile: Affective expressions of machines. *User Modeling and User-Adapted Interaction, 11*(4), 279–295. https://doi.org/10.1023/A:1011811315582

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007). Is the uncanny valley an uncanny cliff? *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication*, 368–373. https://doi.org/10.1109/ROMAN.2007.4415111

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.1996.8.6.551

Biele, C., & Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Experimental Brain Research*. https://doi.org/10.1007/s00221-005-0254-0

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6*(8), 641–651. https://doi.org/10.1038/nrn1724

Calvo, M. G., Avero, P., Fernández-Martín, A., & Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion, 16*(8), 1186–1200. https://doi.org/10.1037/emo0000192

Calvo, M. G., Fernández-Martín, A., Gutiérrez-García, A., & Lundqvist, D. (2018). Selective eye fixations on diagnostic face regions of dynamic emotional expressions: KDEF-dyn database. *Scientific Reports, 8*(1), Article 17039. https://doi.org/10.1038/s41598-018-35259-w

Carlson, C. A., Gronlund, S. D., Weatherford, D. R., & Carlson, M. A. (2012). Processing differences between feature-based facial composites and photos of real faces. *Applied Cognitive Psychology, 26*(4), 525–540. https://doi.org/10.1002/acp.2824

Carter, E. J., Mahler, M., & Hodgins, J. K. (2013). Unpleasantness of animated characters corresponds to increased viewer attention to faces. *Proceedings of the ACM Symposium on Applied Perception*, 35–40. https://doi.org/10.1145/2492494.2502059

Chattopadhyay, D., & MacDorman, K. F. (2016). Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. *Journal of Vision, 16*(11), 1–25. https://doi.org/10.1167/16.11.7

Cheetham, M. (2011). The human likeness dimension of the "uncanny valley hypothesis": Behavioral and functional MRI findings. *Frontiers in Human Neuroscience, 5*(126), 1–14. https://doi.org/10.3389/fnhum.2011.00126

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., & Jancke, L. (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: Eye-tracking data. *Frontiers in Psychology, 4*(108), 1–12. https://doi.org/10.3389/fpsyg.2013.00108

Cheetham, M., Suter, P., & Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the uncanny valley: More like a "happy valley. *Frontiers in Psychology, 5*(1219), 1–15. https://doi.org/10.3389/fpsyg.2014.01219

Cheetham, M., Wu, L., Pauli, P., & Jancke, L. (2015). Arousal, valence, and the uncanny valley: Psychophysiological and self-report findings. *Frontiers in Psychology, 6*(981), 1–15. https://doi.org/10.3389/fpsyg.2015.00981

Chen, F., Wang, L., Peng, G., Yan, N., & Pan, X. (2019). Development and evaluation of a 3-D virtual pronunciation tutor for children with autism spectrum disorders. *PLoS One, 14*(1), 1–22. https://doi.org/10.1371/journal.pone.0210858

Christensen, J. L., Miller, L. C., Appleby, P. R., Corsbie-Massay, C., Godoy, C. G., Marsella, S. C., & Read, S. J. (2013). Reducing shame in a game that predicts HIV risk reduction for young adult MSM: A randomized trial delivered nationally over the web. *Journal of the International AIDS Society, 16*. https://doi.org/10.7448/IAS.16.3.18716

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, M., van Rijn, B., & Chryssafidou, E. (2019). Avatar-based counselling for psychological distress in secondary school pupils: Pilot evaluation. *British Journal of Guidance and Counselling, 47*(4), 446–459. https://doi.org/10.1080/03069885.2018.1506567

Costantini, E., Pianesi, F., & Cosi, P. (2004). Evaluation of synthetic faces: Human recognition of emotional facial displays. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 276–287. https://doi.org/10.1007/978-3-540-24842-2_29

Costantini, E., Pianesi, F., & Prete, M. (2005). Recognising emotions in human and synthetic faces: The role of the upper and lower parts of the face. *Proceedings of the 10th International Conference on Intelligent User Interfaces*, 20–27. https://doi.org/10.1145/1040830.1040846

Craft, A. J. (2012). Love 2.0: A quantitative exploration of sex and relationships in the virtual world second life. *Archives of Sexual Behavior, 41*(4), 939–947. https://doi.org/10.1007/s10508-012-9933-7

Craig, B. M., Mallan, K. M., & Lipp, O. V. (2012). The effect of poser race on the happy categorization advantage depends on stimulus type, set size, and presentation duration. *Emotion, 12*(6), 1303–1314. https://doi.org/10.1037/a0028622

Crookes, K., Ewing, L., Gildenhuys, J. D., Kloth, N., Hayward, W. G., Oxner, M., Pond, S., & Rhodes, G. (2015). How well do computer-generated faces tap face expertise? *PLoS One, 10*(11), 1–18. https://doi.org/10.1371/journal.pone.0141353

Cruwys, T., Alexander Haslam, S., Dingle, G. A., Jetten, J., Hornsey, M. J., Desdemona Chong, E. M., & Oei, T. P. S. (2014). Feeling connected again: Interventions that increase social identification reduce depression symptoms in community and clinical settings. *Journal of Affective Disorders, 159*, 139–146. https://doi.org/10.1016/j.jad.2014.02.019

Dai, Z., & MacDorman, K. F. (2018). The doctor's digital double: How warmth, competence, and animation promote adherence intention. *PeerJ Computer Science, 4*(11), 1–29. https://doi.org/10.7717/peerj-cs.168

Dawel, A., Miller, E. J., Horsburgh, A., & Ford, P. (2021). A systematic survey of face stimuli used in psychological research 2000–2020. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01705-3

Deffler, S. A., Brown, A. S., & Marsh, E. J. (2015). Judging the familiarity of strangers: Does the context matter? *Psychonomic Bulletin & Review, 22*(4), 1041–1047. https://doi.org/10.3758/s13423-014-0769-0

Dellazizzo, L., Potvin, S., Phraxayavong, K., Lalonde, P., & Dumais, A. (2018). Avatar therapy for persistent auditory verbal hallucinations in an ultra-resistant schizophrenia patient: A case report. *Frontiers in Psychiatry, 9*(131), 1–7. https://doi.org/10.3389/fpsyt.2018.00131

Demos, K. E., Kelley, W. M., Ryan, S. L., Davis, F. C., & Whalen, P. J. (2008). Human amygdala sensitivity to the pupil size of others. *Cerebral Cortex, 18*(12), 2729–2734. https://doi.org/10.1093/cercor/bhn034

Diel, A., Weigelt, S., & Macdorman, K. F. (2021). A meta-analysis of the uncanny valley's independent and dependent variables. *ACM Transactions Human-Robot Interaction, 11*(1), 1–33. https://doi.org/10.1145/3470742

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology, 19*, 643–647. https://doi.org/10.1111/j.1469-8986.1982.tb02516.x

Dyck, M., Winbeck, M., Leiberg, S., Chen, Y., Gur, R. C., & Mathiak, K. (2008). Recognition profile of emotions in natural and virtual faces. *PLoS One, 3*(11), 1–8. https://doi.org/10.1371/journal.pone.0003628

Dyck, M., Winbeck, M., Leiberg, S., Chen, Y., & Mathiak, K. (2010). Virtual faces as a tool to study emotion recognition deficits in schizophrenia. *Psychiatry Research, 179*(3), 247–252. https://doi.org/10.1016/j.psychres.2009.11.004

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science, 3*(1), 34–38. https://doi.org/10.1111/j.1467-9280.1992.tb00253.x

Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management, 55*, 62–73. https://doi.org/10.1016/j.tourman.2016.01.013

Fabri, M., Moore, D., & Hobbs, D. (2004). Mediating the expression of emotion in educational collaborative virtual environments: An experimental study. *Virtual Reality, 7*(2), 66–81. https://doi.org/10.1007/s10055-003-0116-7

Fan, S., Ng, T. T., Herberg, J. S., Koenig, B. L., & Xin, S. (2012). Real or fake? Human judgments about photographs and computer-generated images of faces. *SIGGRAPH Asia 2012 Technical Briefs, 1*(212), 3–6. https://doi.org/10.1145/2407746.2407763

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. In *Psychotherapy* (pp. 316–340). Educational Publishing Foundation. https://doi.org/10.1037/pst0000172.

Foo, Y. Z., Nakagawa, S., Rhodes, G., & Simmons, L. W. (2017). The effects of sex hormones on immune function: A meta-analysis. *Biological Reviews, 92*(1), 551–571. https://doi.org/10.1111/brv.12243

Foo, Y. Z., Sutherland, C. A. M., Burton, N. S., Nakagawa, S., & Rhodes, G. (2021). Accuracy in facial trustworthiness impressions: Kernel of truth or modern physiognomy? A meta-analysis. *Personality and Social Psychology Bulletin, 1–17*. https://doi.org/10.1177/01461672211048110

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General, 137*(4), 673–690. https://doi.org/10.1037/a0013875

Freeman, J. B., Johnson, K. L., Ambady, N., & Rule, N. O. (2010a). Sexual orientation perception involves gendered facial cues. *Personality and Social Psychology Bulletin, 36*(10), 1318–1331. https://doi.org/10.1177/0146167210378755

Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010b). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology, 46*(1), 179–185. https://doi.org/10.1016/j.jesp.2009.10.002

Freeman, J. B., Stolier, R. M., Ingbretsen, Z. A., & Hehman, E. A. (2014). Amygdala responsivity to high-level social information from unseen faces. *Journal of Neuroscience, 34*(32), 10573–10581. https://doi.org/10.1523/JNEUROSCI.5063-13.2014

Gaither, S. E., Chen, J. M., Pauker, K., & Sommers, S. R. (2019). At face value: Psychological outcomes differ for real vs. computer-generated multiracial faces. *The Journal of Social Psychology, 159*(5), 592–610. https://doi.org/10.1080/00224545.2018.1538929

Geiger, A. R., & Balas, B. (2021). Robot faces elicit responses intermediate to human faces and objects at face-sensitive ERP components. *Scientific Reports, 11*(1), 1–13. https://doi.org/10.1038/s41598-021-97527-6

Gendron, M., Mesquita, B., & Barrett, L. F. (2013). Emotion perception: Putting the face in context. In *The Oxford handbook of cognitive psychology* (pp. 539–556). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195376746.013.0034.

Gibert, G., Leung, Y., & Stevens, C. J. (2013). Control of speech-related facial movements of an avatar from video. *Speech Communication, 55*(1), 135–146. https://doi.org/10.1016/j.specom.2012.07.001

Gómez-Leal, R., Costa, A., Megías-Robles, A., Fernández-Berrocal, P., & Faria, L. (2021). Relationship between emotional intelligence and empathy towards humans and animals. *PeerJ, 9*, Article e11274. https://doi.org/10.7717/peerj.11274

Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior, 24*(4), 1494–1509. https://doi.org/10.1016/j.chb.2007.05.007

Gonzalez-Franco, M., Bellido, A. I., Blom, K. J., Slater, M., & Rodriguez-Fornells, A. (2016). The neurological traces of look-alike avatars. *Frontiers in Human Neuroscience, 10*(392), 1–10. https://doi.org/10.3389/fnhum.2016.00392

Gračanin, A., Krahmer, E., Balsters, M., Küster, D., & Vingerhoets, A. J. J. M. (2021). How weeping influences the perception of facial expressions: The signal value of tears. *Journal of Nonverbal Behavior, 45*(1), 83–105. https://doi.org/10.1007/s10919-020-00347-x

Green, R. D., MacDorman, K. F., Ho, C. C., & Vasudevan, S. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior, 24*(5), 2456–2474. https://doi.org/10.1016/j.chb.2008.02.019

Guise, V., Chambers, M., Conradi, E., Kavia, S., & Välimäki, M. (2012). Development, implementation and initial evaluation of narrative virtual patients for use in vocational mental health nurse training. *Nurse Education Today, 32*(6), 683–689. https://doi.org/10.1016/j.nedt.2011.09.004

Gwinn, O. S., Matera, C. N., O'Neil, S. F., & Webster, M. A. (2018). Asymmetric neural responses for facial expressions and anti-expressions. *Neuropsychologia, 119*, 405–416. https://doi.org/10.1016/j.neuropsychologia.2018.09.001

Hernandez, N., Metzger, A., Magné, R., Bonnet-Brilhault, F., Roux, S., Barthelemy, C., & Martineau, J. (2009). Exploration of core features of a human face by healthy and autistic adults analyzed by visual scanning. *Neuropsychologia, 47*(4), 1004–1012. https://doi.org/10.1016/j.neuropsychologia.2008.10.023

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2013). Same faces, different labels: Generating the cross-race effect in face memory with social category information. *Memory & Cognition, 41*(7), 1021–1031. https://doi.org/10.3758/s13421-013-0316-7

Hyde, J., Kiesler, S., Hodgins, J. K., & Carter, E. J. (2014). Conversing with children: Cartoon and video people elicit similar conversational behaviors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1787–1796. https://doi.org/10.1145/2556288.2557280

Itz, M. L., Schweinberger, S. R., Schulz, C., & Kaufmann, J. M. (2014). Neural correlates of facilitations in face learning by selective caricaturing of facial shape or reflectance. *NeuroImage, 102*, 736–747. https://doi.org/10.1016/j.neuroimage.2014.08.042

Jackson, P. L., Michon, P. E., Geslin, E., Carignan, M., & Beaudoin, D. (2015). EEVEE: The empathy-enhancing virtual evolving environment. *Frontiers in Human Neuroscience, 9*, 1–16. https://doi.org/10.3389/fnhum.2015.00112

Javor, A., Ransmayr, G., Struhal, W., & Riedl, R. (2016). Parkinson patients' initial trust in avatars: Theory and evidence. *PLoS One, 11*(11), 1–21. https://doi.org/10.1371/journal.pone.0165998

John, L. K., Acquisti, A., & Loewenstein, G. (2011). Strangers on a plane: Context-dependent willingness to divulge sensitive information. *Journal of Consumer Research, 37*(5), 858–873. https://doi.org/10.1086/656423

Joyal, C. C., Jacob, L., Cigna, M. H., Guay, J. P., & Renaud, P. (2014). Virtual faces expressing emotions: An initial concomitant and construct validity study. *Frontiers in Human Neuroscience, 8*(SEP), 1–6. https://doi.org/10.3389/fnhum.2014.00787

Kala, S., Rolison, M. J., Trevisan, D. A., Naples, A. J., Pelphrey, K., Ventola, P., & McPartland, J. C. (2021). Brief report: Preliminary evidence of the N170 as a biomarker of response to treatment in autism spectrum disorder. *Frontiers in Psychiatry, 12*, 1–7. https://doi.org/10.3389/fpsyt.2021.709382

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.17-11-04302.1997

Karras, T., Laine, S., & Aila, T. (2021). A style-based generator architecture for generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*, 4217–4228. https://doi.org/10.1109/TPAMI.2020.2970919

Kätsyri, J. (2018). Those virtual people all look the same to me: Computer-rendered faces elicit a higher false alarm rate than real human faces in a recognition memory task. *Frontiers in Psychology, 9*, 1–12. https://doi.org/10.3389/fpsyg.2018.01362

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*, 1–12. https://doi.org/10.1016/j.neuroimage.2019.116216

Kätsyri, J., de Gelder, B., & Takala, T. (2019). Virtual faces evoke only a weak uncanny valley effect: An empirical investigation with controlled virtual face images. *Perception, 48*(10), 968–991. https://doi.org/10.1177/0301006619869134

Kätsyri, J., Klucharev, V., Frydrych, M., & Sams, M. (2003). Identification of synthetic and natural emotional facial expressions. *Proceedings of the International Conference on Audio-Visual Speech Processing*, 1–5.

Kegel, L. C., Brugger, P., Frühholz, S., Grunwald, T., Hilfiker, P., Kohnen, O., Loertscher, M. L., Mersch, D., Rey, A., Sollfrank, T., Steiger, B. K., Sternagel, J., Weber, M., & Jokeit, H. (2020). Dynamic human and avatar facial expressions elicit differential brain responses. *Social Cognitive and Affective Neuroscience, 15*(3), 303–317. https://doi.org/10.1093/scan/nsaa039

Krumhuber, E. G., Swiderska, A., Tsankova, E., Kamble, S. V., & Kappas, A. (2015). Real or artificial? Intergroup biases in mind perception in a cross-cultural perspective. *PLoS One, 10*(9), 1–14. https://doi.org/10.1371/journal.pone.0137840

Krumhuber, E. G., Tamarit, L., Roesch, E. B., & Scherer, K. R. (2012). FACSGen 2.0 animation software: Generating three-dimensional FACS-valid facial expressions for emotion research. *Emotion, 12*(2), 351–363. https://doi.org/10.1037/a0026632

Lewkowicz, D. J., & Ghazanfar, A. A. (2012). The development of the uncanny valley in infants. *Developmental Psychobiology, 54*(2), 124–132. https://doi.org/10.1002/dev.20583

Lidestam, B., Lyxell, B., & Lundeberg, M. (2001). Speech-reading of synthetic and natural faces: Effects of contextual cueing and mode of presentation. *Scandinavian Audiology, 30*(2), 89–94. https://doi.org/10.1080/010503901300112194

Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2006.09.002

Lucas, G. M., Gratch, J., King, A., & Morency, L. P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior, 37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., & Morency, L. P. (2017). Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers. *Frontiers Robotics AI, 4*(51), 1–9. https://doi.org/10.3389/frobt.2017.00051

MacDorman, K. F. (2019). In the uncanny valley, transportation predicts narrative enjoyment more than empathy, but only for the tragic hero. *Computers in Human Behavior, 94*, 140–153. https://doi.org/10.1016/j.chb.2019.01.011

MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition, 146*, 190–205. https://doi.org/10.1016/j.cognition.2015.09.019

MacDorman, K. F., & Chattopadhyay, D. (2017). Categorization-based stranger avoidance does not explain the uncanny valley effect. *Cognition, 161*, 132–135. https://doi.org/10.1016/j.cognition.2017.01.009

MacDorman, K. F., Green, R. D., Ho, C. C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25*(3), 695–710. https://doi.org/10.1016/j.chb.2008.12.026

Malek, N., Messinger, D., Gao, A. Y. L., Krumhuber, E., Mattson, W., Joober, R., Tabbane, K., & Martinez-Trujillo, J. C. (2019). Generalizing Duchenne to sad expressions with binocular rivalry and perception ratings. *Emotion, 19*(2), 234–241. https://doi.org/10.1037/emo0000410

Maras, M.-H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence and Proof, 23*(3), 255–262. https://doi.org/10.1177/1365712718807226

Matheson, H. E., Bilsbury, T. G., & McMullen, P. A. (2012). Second-order relational face processing is applied to faces of different race and photographic contrast. *Canadian Journal of Experimental Psychology, 66*(1), 51–62. https://doi.org/10.1037/a0026062

Matheson, H. E., & McMullen, P. A. (2011). A computer-generated face database with ratings on realism, masculinity, race, and stereotypy. *Behavior Research Methods, 43*, 224–228. https://doi.org/10.3758/s13428-010-0029-9

Matheson, H. E., & McMullen, P. A. (2011). A computer-generated face database with ratings on realism, masculinity, race, and stereotypy. *Behavior Research Methods, 43*(1), 224–228. https://doi.org/10.3758/s13428-010-0029-9

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences, 11*(1), 8–15. https://doi.org/10.1016/j.tics.2006.11.002
McKone, E., Wan, L., Pidcock, M., Crookes, K., Reynolds, K., Dawel, A., Kidd, E., & Fiorentini, C. (2019). A critical period for faces: Other-race face recognition is improved by childhood but not adult social contact. *Scientific Reports, 9*(1), 1–13. https://doi.org/10.1038/s41598-019-49202-0
Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3–35. https://doi.org/10.1037//1076-8971.7.1.3
Milcent, A. S., Geslin, E., Kadri, A., & Richir, S. (2019). Expressive virtual human: Impact of expressive wrinkles and pupillary size on emotion recognition. *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 215–217. https://doi.org/10.1145/3308532.3329446
Miller, E. J., Krumhuber, E. G., & Dawel, A. (2020). Observers perceive the Duchenne marker as signaling only intensity for sad expressions, not genuine emotion. *Emotion*. https://doi.org/10.1037/emo0000772
Mori, M. (1970). Bukimi no tani [The uncanny valley]. Energy, 7(4), 33-35.
Moser, E., Derntl, B., Robinson, S., Fink, B., Gur, R. C., & Grammer, K. (2007). Amygdala activation at 3T in response to human and avatar facial expressions of emotions. *Journal of Neuroscience Methods, 161*(1), 126–133. https://doi.org/10.1016/j.jneumeth.2006.10.016
Mühlberger, A., Wieser, M. J., Herrmann, M. J., Weyers, P., Tröger, C., & Pauli, P. (2009). Early cortical processing of natural and artificial emotional faces differs between lower and higher socially anxious persons. *Journal of Neural Transmission, 116*(6), 735–746. https://doi.org/10.1007/s00702-008-0108-6
Mundy, M. E., Downing, P. E., & Graham, K. S. (2012). Extrastriate cortex and medial temporal lobe regions respond differentially to visual feature overlap within preferred stimulus category. *Neuropsychologia, 50*(13), 3053–3061. https://doi.org/10.1016/j.neuropsychologia.2012.07.006
Mustafa, M., Guthe, S., Tauscher, J. P., Goesele, M., & Magnor, M. (2017). How human am I? EEG-based evaluation of animated virtual characters. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 5098–5108. https://doi.org/10.1145/3025453.3026043
Mustafa, M., & Magnor, M. (2016). EEG based analysis of the perception of computer-generated faces. *Proceedings of the 13th European Conference on Visual Media Production*, 1–10. https://doi.org/10.1145/2998559.2998563
Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. In *Biological reviews*. https://doi.org/10.1111/j.1469-185X.2007.00027.x
Naples, A., Nguyen-Phuc, A., Coffman, M., Kresse, A., Faja, S., Bernier, R., & McPartland, J. C. (2015). A computer-generated animated face stimulus set for psychophysiological research. *Behavior Research Methods, 47*, 562–570. https://doi.org/10.3758/s13428-014-0491-x
Ni, H., Wang, J., Wang, L., & Yan, N. (2018). Track your emotional perception of 3-D virtual talking head in human-computer interaction. *IEEE International Conference on Cyborg and Bionic Systems*, 298–303. https://doi.org/10.1109/CBS.2018.8612271
Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences of the United States of America, 119*(8), 2–4. 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 of the United States of America*. https://doi.org/10.1073/pnas.0805664105
Papesh, M. H., & Goldinger, S. D. (2009). Deficits in other-race face recognition: No evidence for encoding-based effects. *Canadian Journal of Experimental Psychology, 63*(4), 253–262. https://doi.org/10.1037/a0015802
Parmar, D., Olafsson, S., Utami, D., Murali, P., & Bickmore, T. (2022). Designing empathic virtual agents: Manipulating animation, voice, rendering, and empathy to create persuasive agents. *Autonomous Agents and Multi-Agent Systems, 36*(1), 17. https://doi.org/10.1007/s10458-021-09539-1
Patel, H., & MacDorman, K. F. (2015). Sending an avatar to do a human's job: Compliance with authority persists despite the uncanny valley. *Presence, 24*(1), 1–23. https://doi.org/10.1162/PRES_a_00212
Perizzolo Pointet, V. C., Moser, D. A., Vital, M., Rusconi Serpa, S., Todorov, A., & Schechter, D. S. (2020). Violence exposure is associated with atypical appraisal of threat among women: An EEG study. *Frontiers in Psychology, 11*, 1–18. https://doi.org/10.3389/fpsyg.2020.576852
Peromaa, T., & Olkkonen, M. (2019). Red color facilitates the detection of facial anger — but how much? *PLoS One, 14*(4), 1–13. https://doi.org/10.1371/journal.pone.0215610
Philip, L., Martin, J.-C., & Clavel, C. (2018). Rapid facial reactions in response to facial expressions of emotion displayed by real versus virtual faces. *i-Perception, 9*(4), 1–18. https://doi.org/10.1177/2041669518786527.
Pickard, M. D., Roster, C. A., & Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior, 65*, 23–30. https://doi.org/10.1016/j.chb.2016.08.004
Recio, G., Sommer, W., & Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Research, 1376*, 66–75. https://doi.org/10.1016/j.brainres.2010.12.041
Riedl, R., Mohr, P., Kenning, P., Davis, F., & Heekeren, H. (2014). Trusting humans and avatars: A brain imaging study based on evolution theory. *Journal of Management Information Systems, 30*(4), 83–114. https://doi.org/10.2753/MIS0742-1222300404
Rohatgi, A. (2021). WebPlotDigitizer. https://automeris.io/WebPlotDigitizer.
Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica, 128*(2), 274–289. https://doi.org/10.1016/j.actpsy.2008.02.003
Rossion, B., & Jacques, C. (2012). The N170: Understanding the time course of face perception in the human brain. In *The Oxford handbook of event-related potential components* (pp. 115–141). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0064.
Roth, D., Bloch, C., Schmitt, J., Frischlich, L., Latoschik, M. E., & Bente, G. (2019). Perceived authenticity, empathy, and pro-social intentions evoked through avatar-mediated self-disclosures. *Proceedings of Mensch Und Computer*, 21–30. https://doi.org/10.1145/3340764.3340797
Royer, J., Blais, C., Charbonneau, I., Déry, K., Tardif, J., Duchaine, B., Gosselin, F., & Fiset, D. (2018). Greater reliance on the eye region predicts better face recognition ability. *Cognition, 181*, 12–20. https://doi.org/10.1016/j.cognition.2018.08.004
Rubin, A., Livingston, N. A., Brady, J., Hocking, E., Bickmore, T., Sawdy, M., Kressin, N., Saitz, R., & Simon, S. (2022). Computerized relational agent to deliver alcohol brief intervention and referral to treatment in primary care: A randomized clinical trial. *Journal of General Internal Medicine, 30*, 70–77. https://doi.org/10.1007/s11606-021-06945-9
de Ruiter, A. (2021). The distinct wrong of deepfakes. *Philosophy & Technology, 34*(4), 1311–1332. https://doi.org/10.1007/s13347-021-00459-2
Schindler, S., Zell, E., Botsch, M., & Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Scientific Reports, 7*, 1–13. https://doi.org/10.1038/srep45003
Schyns, P. G., Jentzsch, I., Johnson, M., Schweinberger, S. R., & Gosselin, F. (2003). A principled method for determining the functionality of brain responses. *NeuroReport, 14*(13), 1665–1669. https://doi.org/10.1097/00001756-200309150-00002
Seo, Y., Kim, M., Jung, Y., & Lee, D. (2017). Avatar face recognition and self-presence. *Computers in Human Behavior, 69*, 120–127. https://doi.org/10.1016/j.chb.2016.12.020
Shen, B., RichardWebster, B., O'Toole, A., Bowyer, K., & Scheirer, W. J. (2021). A study of the human perception of synthetic faces. *Proceedings of the 16th IEEE International Conference on Automatic Face and Gesture Recognition*, 1–8. https://doi.org/10.1109/FG52635.2021.9667066
Shepherd, R.-M., & Edelmann, R. J. (2005). Reasons for internet use and social anxiety. *Personality and Individual Differences, 39*(5), 949–958. https://doi.org/10.1016/j.paid.2005.04.001
Singh, B., Mellinger, C., Earls, H. A., Tran, J., Bardsley, B., & Correll, J. (2021). Does cross-race contact improve cross-race face perception? A meta-analysis of the cross-race deficit and contact. *Personality and Social Psychology Bulletin, 48*(6), 865–887. https://doi.org/10.1177/01461672211024463
Sollfrank, T., Kohnen, O., Hilfiker, P., Kegel, L. C., Jokeit, H., Brugger, P., Loertscher, M. L., Rey, A., Mersch, D., Sternagel, J., Weber, M., & Grunwald, T. (2021). The effects of dynamic and static emotional facial expressions of humans and their avatars on the EEG: An ERP and ERD/ERS study. *Frontiers in Neuroscience, 15*, 1–14. https://doi.org/10.3389/fnins.2021.651044
Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*. https://doi.org/10.1016/j.cognition.2012.12.001
Syrjämäki, A. H., Isokoski, P., Surakka, V., Pasanen, T. P., & Hietanen, J. K. (2020). Eye contact in virtual reality – a psychophysiological study. *Computers in Human Behavior, 112*, 1–10. https://doi.org/10.1016/j.chb.2020.106454
Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 46A*(2), 225–245. https://doi.org/10.1080/14640749308401045
Thorstenson, C. A., Pazda, A. D., Young, S. G., & Elliot, A. J. (2019). Face color facilitates the disambiguation of confusing emotion expressions: Toward a social functional account of face color in emotion communication. *Emotion, 19*(5), 799–807. https://doi.org/10.1037/emo0000485
Tinwell, A., Grimshaw, M., & Nabi, D. A. (2015). The effect of onset asynchrony in audio-visual speech and the Uncanny Valley in virtual characters. *International Journal of Mechanisms and Robotic Systems, 2*(2), 97. https://doi.org/10.1504/ijmrs.2015.068991
Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior, 27*(2), 741–749. https://doi.org/10.1016/j.chb.2010.10.018
Tinwell, A., Nabi, D. A., & Charlton, J. P. (2013). Perception of psychopathy and the Uncanny Valley in virtual characters. *Computers in Human Behavior, 29*(4), 1617–1625. https://doi.org/10.1016/j.chb.2013.01.008
Tinwell, A., & Sloan, R. J. S. (2014). Children's perception of uncanny human-like virtual characters. *Computers in Human Behavior, 36*, 286–296. https://doi.org/10.1016/j.chb.2014.03.073
Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589.
Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2008.10.001
Vaitonytė, J., Blomsma, P. A., Alimardani, M., & Louwerse, M. M. (2021). Realism of the face lies in skin and eyes: Evidence from virtual and human agents. *Computers in Human Behavior Reports, 3*, 1–16. https://doi.org/10.1016/j.chbr.2021.100065

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 43*(2), 161–204. https://doi.org/10.1080/14640749108400966

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *Quarterly Journal of Experimental Psychology, 69*(10), 1996–2019. https://doi.org/10.1080/17470218.2014.990392

Vernon, R., Sutherland, C. A. M., Young, A., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences, 111*(32), E3353–E3361. https://doi.org/10.1073/pnas.1409860111

Viechtbauer, W. (2010). Conducting Meta-analysis in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03.

Wang, J., Chen, J., Yan, N., Wang, L., & Ng, L. (2019). How do we perceive emotion from a 3-D virtual talking head: Evidence from functional near-infrared spectroscopy. *International Conference On Interfaces and Human Computer Interaction*, 115–122. https://doi.org/10.33965/ihci2019_201906l015

Wang, L. C., & Fodness, D. (2010). Can avatars enhance consumer trust and emotion in online retail sales? *International Journal of Electronic Marketing and Retailing, 3*(4). https://doi.org/10.1504/IJEMR.2010.036881, 341–36.

Wang, N., Xu, S., Zhang, S., Luo, Y., & Geng, H. (2019). ERP evidence on how gaze convergence affects social attention. *Scientific Reports, 9*(1), 1–11. https://doi.org/10.1038/s41598-019-44058-w

Wieser, M. J., & Brosch, T. (2012). Faces in context: A review and systematization of contextual influences on affective face processing. *Frontiers in Psychology, 3*, 471. https://doi.org/10.3389/fpsyg.2012.00471

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

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*. https://doi.org/10.1177/0956797615590992

Yin, R. K. (1969). Looking at upide-down faces. *Journal of Experimental Psychology, 81*(1), 141–145. https://doi.org/10.1037/h0027474

Young, A. W., Hellawell, D., & Hay, D. C. (2013). Configurational information in face perception. *Perception, 16*(6), 747–759. https://doi.org/10.1068/p160747n

Zhao, T., Chen, J., Wang, L., & Yan, N. (2020). Emotional appraisal processing of computer-generated facial expressions: An functional near-infrared spectroscopy study. *NeuroReport, 31*(6), 437–441. https://doi.org/10.1097/WNR.0000000000001420

Zhou, Y., Chen, F., Chen, H., Wang, L., & Yan, N. (2016). Evaluation of a multimodal 3-D pronunciation tutor for learning Mandarin as a second language: An eye-tracking study. *Proceedings of 2016 10th International Symposium on Chinese Spoken Language Processing, 1–5*. https://doi.org/10.1109/ISCSLP.2016.7918451