



A systematic survey of face stimuli used in psychological research 2000–2020

Amy Dawel¹ · Elizabeth J. Miller¹ · Annabel Horsburgh¹ · Patrice Ford¹

Accepted: 7 September 2021 / Published online: 3 November 2021
© The Psychonomic Society, Inc. 2021

Abstract

For decades, psychology has relied on highly standardized images to understand how people respond to faces. Many of these stimuli are rigorously generated and supported by excellent normative data; as such, they have played an important role in the development of face science. However, there is now clear evidence that testing with ambient images (i.e., naturalistic images “in the wild”) and including expressions that are spontaneous can lead to new and important insights. To precisely quantify the extent to which our current knowledge base has relied on standardized and posed stimuli, we systematically surveyed the face stimuli used in 12 key journals in this field across 2000–2020 ($N = 3374$ articles). Although a small number of posed expression databases continue to dominate the literature, the use of spontaneous expressions seems to be increasing. However, there has been no increase in the use of ambient or dynamic stimuli over time. The vast majority of articles have used highly standardized and nonmoving pictures of faces. An emerging trend is that virtual faces are being used as stand-ins for human faces in research. Overall, the results of the present survey highlight that there has been a significant imbalance in favor of standardized face stimuli. We argue that psychology would benefit from a more balanced approach because ambient and spontaneous stimuli have much to offer. We advocate a cognitive ethological approach that involves studying face processing in natural settings as well as the lab, incorporating more stimuli from “the wild”.

Keywords Facial expression · Spontaneous · Posed · Ambient · Virtual avatar

Faces are a key topic for many sub-disciplines of psychology because of their centrality to understanding human interpersonal behavior (e.g., clinical, cognitive, social, and developmental psychology; also of interest in neuroscience and artificial intelligence). Across psychology, there has been a longstanding tradition of using highly standardized face images. Face stimuli are typically shown under controlled lighting conditions, in frontal view, or a small number of other standard viewpoints. Some studies take this standardization a step further by editing out hair and identifying marks (e.g., moles). Additionally, emotional face

images (e.g., happy, sad) have been generated primarily by asking models to pose expressions, and are often perceived as faking emotion (Dawel et al., 2017). Overall, this approach to stimulus development gives excellent experimental control and has created a rigorous empirical base. However, an unfortunate side effect is that we have limited knowledge about how people respond to the full repertoire of faces and facial behavior we see in real life. Major reviews over many years have highlighted that the low use of spontaneous expressions is a serious problem for the emotion literature (e.g., Barrett et al., 2019; Russell & Fernández-Dols, 1997). More recently, other areas of face research have discovered that using ambient in addition to standardized images can provide important new insights into face processing (Burton, 2013; Jenkins et al., 2011; Sutherland et al., 2013). Here, we briefly review key evidence that shows that using ambient as well as standardized face images, and spontaneous as well as posed expressions, is critical for moving face science forward. We then survey the use of face stimuli in the psychological and neuroscience literature since the turn of the century. Our study aims to build momentum towards a more balanced approach by providing the first empirical data to quantify trends in the use of standardized and posed face stimuli relative to ambient and spontaneous ones. We also track changes in the total number of articles using faces as

✉ Amy Dawel
amy.dawel@anu.edu.au

Elizabeth J. Miller
elizabeth.miller@anu.edu.au

Annabel Horsburgh
annabel.horsburgh@gmail.com

Patrice Ford
patrice.a.ford@gmail.com

¹ Research School of Psychology (building 39), The Australian National University, Canberra ACT 2600, Australia

stimuli, and the use of dynamic (moving) expressions and virtual face stimuli.

New insights from ambient face stimuli

The term ambient is used in the face literature to refer to images that are captured under naturally occurring conditions, “in the wild” (e.g., Sutherland et al., 2013). For instance, if you were to search Google Images for Cate Blanchett, it would produce a range of images showing Cate from different viewpoints, under different lighting conditions, and with different expressions and hairstyles. Even though these images all show Cate, they may look quite different. This variability has been critical to gaining a new perspective on face processing over the past decade.

In one key example, testing with ambient images has revealed important differences between familiar and unfamiliar face processing. It turns out that, while humans are very good at connecting different photographic or retinal images of a familiar person together, they are very poor at doing so for unfamiliar faces. Jenkins et al. (2011) asked participants to group 40 ambient photographs of two familiar or two unfamiliar identities into same-identity piles. For familiar faces, participants grouped the photographs into two piles—one for each identity—almost perfectly. In contrast, for unfamiliar faces participants always made more than two piles, with a median of 7.5 piles. These findings suggest that familiar but not unfamiliar face processing is robust to image variation, and show that the unfamiliar face problem is one of “telling faces together” rather than “telling them apart”. Overall, this line of research has highlighted key differences in familiar and unfamiliar face processing (for review, see Burton, 2013) that had previously been overlooked, partly because tests that use standardized face images overestimate how good people are at grouping together unfamiliar faces (Ritchie et al., 2015).

Testing with ambient images has also revealed a new dimension of facial first impressions, not previously evident when standardized images were used. When standardized images were used, facial first impressions appeared to have two key dimensions: trustworthiness and dominance (Oosterhof & Todorov, 2008; Todorov et al., 2008). However, when Sutherland et al. (2013) retested this idea using 1000 ambient images, they found a third dimension emerged: youthfulness/attractiveness. In hindsight, it is apparent that Todorov and colleagues did not find this third dimension because the standardized images they used (photographs from the Karolinska Directed Emotional Faces [KDEF] set from Lundqvist et al., 1998; and computer-generated FaceGen images) did not vary much in ways that are relevant to this dimension. For instance, all of the KDEF models are young adults. While it is possible that the youthfulness/attractiveness dimension could have been found with another set of standardized faces that varied in age, standardized databases invariably eliminate important real-world information that can create this kind of sampling error. Overall, these two examples highlight that relying on standardized stimuli alone is not just a problem of ecological validity. Using only stimuli that eliminate important real-world information may lead to missing or misunderstanding key aspects of human psychological processes, ultimately leading theorizing astray.

Posed and spontaneous facial expressions

For facial expression researchers, there is an additional layer of standardization: most stimuli have been generated by asking people to pose expressions. The focus has been on developing highly standardized images that people agree show particular emotions (e.g., anger, sadness, etc.). Consequently, the displays are intense, with minimal variation across individuals (e.g., “angry” facial configuration for person X is very similar to that for person Y, etc.). While this standardization can be a strength for

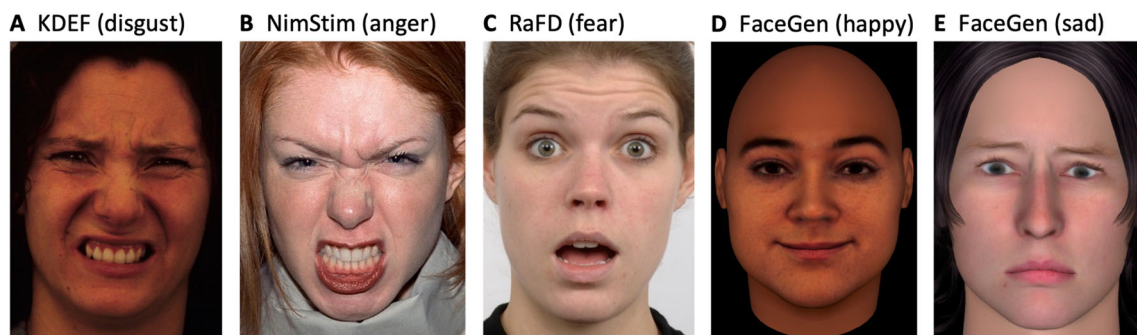


Fig. 1 A–C Example stimuli from popular photographic databases, in which facial expressions are posed. Permission to publish these images in scientific publications is given at: (A) <http://kdef.se/home/using%20and%20publishing%20kdef%20and%20akdef.html> (we show KDEF image AF02DIS); (B) <https://www.macbrain.org/resources.htm> (we show NimStim image 01F_AN_O); and (C) www.socsci.ru.nl:8180/RaFD2/RaFD?p=faq (we show RaFD image Rafd090_32_Caucasian_female_surprized_frontal). Note, we do not include a PoFA stimulus

because of copyright restrictions, but for examples see: <https://www.paulekman.com/product/pictures-of-facial-affect-pofa/>). D, E Example virtual face stimuli made with FaceGen Modeller Pro (Singular Inversions Inc., 2009). D An avatar generated by importing images from the KDEF (identity F03) into FaceGen. E A randomly generated Caucasian female, with facial texture added, showing a sad expression at 1.0 intensity

experimental work, these types of images do not reflect the extensive variation seen in real-world expressions and are often perceived as faking emotion (Dawel et al., 2017).

Adding more spontaneous expressions to the research mix may provide new insights into emotional face processing. For example, studies that have tested people's ability to recognize emotions from spontaneous expressions find performance is considerably worse than for posed expressions (Aviezer et al., 2012, 2015; Naab & Russell, 2007). This line of work suggests that the emotional signals sent by faces are not always easy to read. Another major line of work around smiling shows that people can tell the difference between genuinely felt and posed happiness and vary in their responses to different smile types (e.g., Dawel et al., 2015; Dawel, Dumbleton, et al., 2019a; Gunnery & Ruben, 2016; Heerey, 2014; Shore & Heerey, 2011). There are even instances when using spontaneous expressions alongside posed ones has led to different research conclusions. People with schizophrenia are typically poorer than healthy controls at identifying emotion in faces (for meta-analysis, see Kohler et al., 2010), but this pattern reverses when expressions are spontaneous rather than posed (Davis & Gibson, 2000; LaRusso, 1978), indicating that some types of emotion processing may be relatively intact in schizophrenia. In another example, a major theory of the affective deficits associated with psychopathy (Blair, 2006) was supported when tested with spontaneous expressions, but not when tested with posed ones (Dawel, Wright, et al., 2019b). Overall, this evidence suggests that an approach that balances the use of posed expressions in research with more spontaneous ones has potential to strengthen understanding of emotional processes.

The importance of studying spontaneous expressions was recognized several decades ago (e.g., Russell & Fernández-Dols, 1997). Yet psychology has continued to use mostly posed expressions. One of the reasons for this pattern is that, until recently, researchers had good access to only a small number of facial expression databases—all of which show static, posed expressions in controlled settings. The first database to become widely available was Ekman and Friesens' Pictures of Facial Affect (PoFA; 1976), which is still influential today (e.g., citations = 5680, with 1110 cites since 2017, from Google Scholar 27 May 2021). Most of the PoFA stimuli were created using the Directed Facial Action Task (Ekman, 2007), in which expressers are instructed to pose specific facial actions theoretically associated with basic emotion categories (Ekman, 1992). For example, when posing "anger", expressers are told to pull their eyebrows down and together, raise their upper eyelids, and press their lips together (Levenson et al., 1990). The word "anger" is never mentioned. Similarly, to pose "disgust", expressers are told to wrinkle their nose, let their lips part, and pull their lower lip down, with no mention of "disgust" per se (Ekman, 2007). The one exception is the PoFA happy expressions, which were captured "off guard during a spontaneously occurring happy moment in the photographic session" (Ekman, 1980, p. 834). Ekman and colleagues later generated several related databases, including the Japanese

and Caucasian Facial Expressions of Emotions (JACFEE; Matsumoto & Ekman, 1988), which shows Japanese expressers in addition to Caucasian ones, and the Facial Expressions of Emotion—Stimuli and Tests (FEEST; Young et al., 2002), which shows morphed versions of PoFA stimuli. The present study groups all of Ekman and colleagues' sets together, due to overlapping stimuli belonging to each set and/or the same methods being used to produce the expressions.

More recently, several other expression databases have gained in popularity, including the KDEF (Lundqvist et al., 1998), NimStim (Tottenham et al., 2009), and Radboud Faces Database (RaFD, Langner et al., 2010; see Fig. 1 for example stimuli). All the RaFD expressions, including the happy ones, were elicited using the Directed Facial Action Task (Langner et al., 2010). In contrast, the NimStim and KDEF expressions were elicited by instructing actors to pose named emotions. NimStim expressions were elicited by instructing professional actors to "pose a particular expression (e.g., "make a happy face") and produce the facial expression as they saw fit" (Tottenham et al., 2009, p. 243). For the KDEF, amateur actors read information about the expressions and rehearsed them for an hour before being instructed to "evoke the emotion [to be] expressed, and—while maintaining a way of expressing the emotion that [feels] natural to [you]—try and make the expression strong and clear" (Lundqvist et al., 1998).

In the present work, we anticipated that these four sets—the KDEF, NimStim, RaFD, and Ekman and colleagues'—had probably been used in a large amount of empirical work. We aimed to determine exactly how much they have dominated the literature, what the relative contribution of each database has been, and whether there have been any changes in their contributions over time.

Virtual faces

Technology now allows researchers to easily and affordably create virtual human faces. For example, Fig. 1E illustrates a Caucasian female randomly generated with FaceGen Modeller Pro (Singular Inversions Inc., 2009), a popular software package available to anyone for a reasonable fee (e.g., US\$699 on 13 May 2021, <https://facegen.com/products.htm>). Investigating how humans perceive and respond to virtual faces is important, as avatars and other virtual beings are increasingly becoming part of our daily lives (e.g., in e-therapy, Dellazizzo et al., 2018; online dating, www.ivirtual.com; and medical training, Andrade et al., 2010; Guise et al., 2012). However, our impression is that virtual faces are mostly being used to address questions that are supposedly about how humans respond to *human* faces. Virtual faces have some obvious advantages (e.g., high levels of stimulus control and easy manipulation of experimental parameters). However, the current evidence does not support replacing images of human faces with virtual ones. For example, the

types of virtual faces used to date show reduced other-race effects (Craig et al., 2012; Crookes et al., 2015), are more difficult to remember (Balas & Pacella, 2015), and elicit less neural activation (Katsyri et al., 2019; Schindler et al., 2017) than human faces. In addition, just as testing with ambient images can lead to different outcomes than testing with standardized ones, testing with virtual faces can also produce findings that do not replicate with human faces (e.g., in emotion perception research, Miller et al., 2020).

The present work is the first to assess the extent to which virtual faces are being used in research, including how often they are being used in place of real human faces to answer questions about human responses to human faces. We defined virtual faces as those created with any virtual or avatar face software or generative adversarial networks (GANs; Goodfellow et al., 2014; Karras et al., 2019). Examples of virtual face software include FaceGen (Singular Inversions Inc., 2009; alone or in combination with FACSGen software, which is used to manipulate facial expressions in FaceGen faces; Krumhuber et al., 2012) and Poser, another software package that is popular with researchers (<https://www.posersoftware.com>). In contrast, GANs are produced by machine learning algorithms, which are trained on human photographs to create new, realistic photo-quality images of people who do not exist in the real world (Goodfellow et al., 2014; for examples, see www.thispersondoesnotexist.com).

The present study

Our survey aimed to systematically evaluate recent trends, from 2000 to 2020, in the types of faces being used as stimuli in psychology and adjacent fields. We identified 12 journals that have published large amounts of research on faces across different core topics and methods in psychology and adjacent fields (vision, neuroscience) and screened them for articles that used faces as stimuli. We coded what type of face stimuli they used (photographic images of humans, virtual faces, other human-like faces), where stimuli were sourced from (database name, whether images were standardized or ambient), whether stimuli were static or dynamic, the type of question the article addressed (expression/emotion in the face, or another type of question, e.g., face identity recognition or adaptation effects, other-race face recognition, eye-gaze without emotion, social judgments about faces, etc.) and, if expression/emotion-related, whether expressions were posed or spontaneous. Our reasons for coding whether the type of question each article addressed involved expression/emotion were that: one of our key questions concerned the types of expression stimuli being used in research; and our initial exploratory search indicated that enough papers tested expression/emotion questions for meaningful analyses to be conducted within this subset. Posed expressions were defined as expressions that

were generated using the Directed Facial Action Task (Ekman, 2007) or by instructions to pose or act out an expression, including method acting. Spontaneous expressions were defined as expressions that occurred without any instructions to produce a facial display (e.g., produced during social interactions, or in reaction to some stimulus). We also coded the dependent variables each article used to measure people's responses to face stimuli (see Supplement S9 for these results).

Our analyses addressed five core themes: (1) Are face stimuli being used more or less in research over time? (2) How much has the literature relied on standardized face stimuli relative to ambient ones? (3) How much has face expression/emotion work relied on posed expressions and, more specifically, the most popular posed expression databases? (4) How often are facial expressions presented dynamically, and is this improving? And (5) to what extent are virtual faces being used, including as replacements for real human ones?

METHOD

Scopus selection of journals for survey

Figure 2 outlines the journal article search and screening process, including the number of articles retained and excluded at each stage. Supplement S1 provides the details of our search and justification for our journal selection. We aimed to survey journals that: (1) published substantive amounts of research on face processing; (2) are high-quality; and (3) cover a broad range of psychology and psychology-adjacent topics and methods. To identify the journals that published the largest number of articles on faces since the turn of the century, we conducted an initial search in Scopus for articles that contained the term *face** or *facial* or *eye-gaze* in the title, abstract, or keywords, limited to the disciplines of psychology and neuroscience, from 2000 to 2020 (final update of search performed on 27 May 2021). Initial testing of different search terms found *expression* was almost always accompanied by *face** or *facial** so we omitted this term (i.e., because it was redundant). However, *eye-gaze* sometimes appeared without *face** or *facial** so we retained this term. We initially retained the 37 journals that had published >250 articles (as of 20 August 2020) that met our search criteria. From these, we selected 12 journals that were representative of the Journal Citation Reports topic areas and mean and range of impact factors of the 37 journals (see Supplement S1). The 12 journals we surveyed were (in alphabetical order): *Cognition*, *Cognition and Emotion* (C&E), *Computers in Human Behavior* (CHB), *Emotion* (first published 2001), *Frontiers in Human Neuroscience* (FrHN, first published 2008), *Journal of Affective Disorders* (JAD), *Journal of Personality and Social Psychology* (JPSP), *Journal of Vision* (JoV, first published 2001), *Neuropsychologia* (NPLogia), *Perception*, *Psychological Science* (Psych Sci), and *Social Cognitive and Affective Neuroscience* (SCAN, first published 2006).

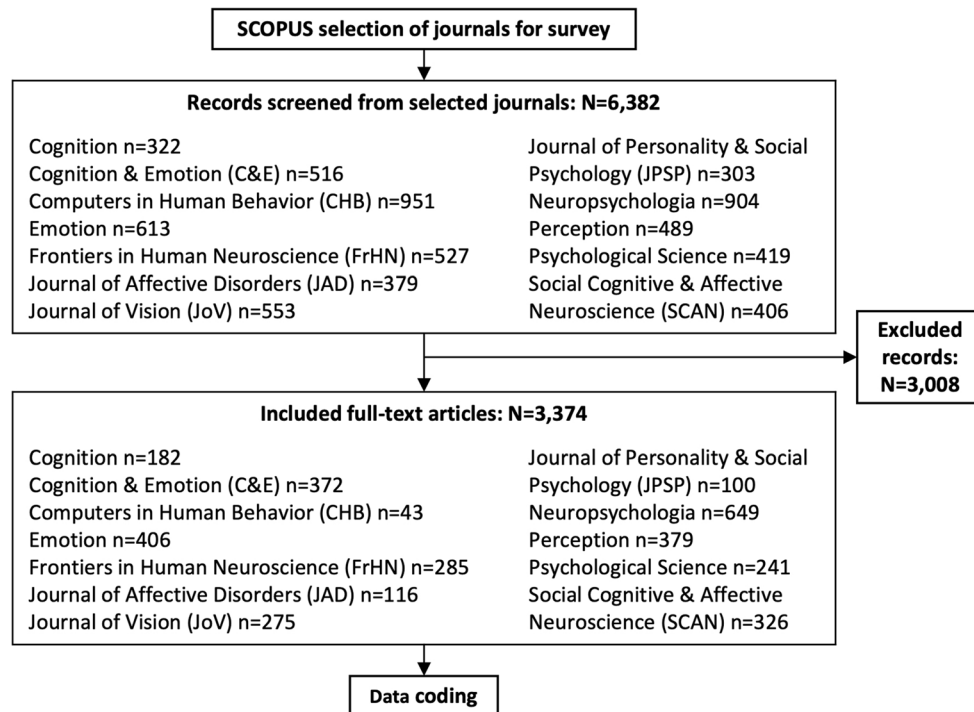


Fig. 2 Flow chart outlining the journal article search and screening process

Article screening: Inclusion and exclusion criteria

To determine whether each article used faces as stimuli, we first screened the abstracts returned by our search. If an abstract did not make this information clear, screening progressed to the article full text. Inclusion/exclusion criteria were as follows: (1) the article must be a full-text article (published conference abstracts were excluded); (2) the article must include empirical work that used faces as stimuli (review articles were excluded); (3) stimuli must depict faces in human form (i.e., could be photographs or videos of humans, or virtual human faces, GAN human faces, or even cartoons or line drawings of human faces, but not animal or robot faces); (4) stimuli must include the entire internal region of faces (e.g., faces with hair cropped out were included, but eyes-only stimuli were excluded); (5) stimulus images could be edited, but must still be recognizable as faces (e.g., morphed and averaged faces were included); and (6) faces must be a stimulus of interest in one or more experiments. Each abstract/article was originally screened by one of the authors or ZW (see acknowledgments). To check the reliability of our screening, an additional coder (WW, see acknowledgments) screened a random sample of $N = 310$ articles. Agreement between the original screening and that of WW was 97.4%.

Data coding

Supplement S2 lists the coding variables and options. Each article was originally coded by one of the authors. Where the standardized/ambient nature of images was not specified in

the text, but example stimuli or other relevant information were available, the coder made a judgment about the nature of the images if it was reasonably possible to do so. For instance, famous faces were typically coded as ambient images, and facial expressions displayed by infants were coded as spontaneous on the basis that an infant could not be verbally directed to pose an expression. If an article referred to another study for stimulus details, that article was also sourced for further information. To check the reliability of our coding, WW coded the $N = 160$ articles he found met inclusion criteria when he screened the random sample of $N = 310$ articles. Agreement between the original coding and that of WW was 95%.

Data analysis

Because of the nature of our data and questions, our results are primarily descriptive. We also present linear or quadratic fits and their associated Pearson's r or R^2 values in text and figures where we track changes across time from 2000 to 2020. The data we describe is of two main types: count data and percentages. Count data indicate the absolute number of articles that met our study inclusion criteria, split by different combinations of coding variables—for example, the number of articles that were published in a given journal in a given year. Count data are used to describe the patterns we observed in absolute terms. For example, whether there was any change in the absolute number of articles that used faces as stimuli across 2000 to 2020. In contrast, percentages are used to describe

patterns in the use of particular types of face stimuli in relation to overall changes in publishing. For example, we wanted to know whether any change we observed across 2000 to 2020 in the number of articles that used spontaneous or dynamic expressions was simply tracking broader changes in the number of articles testing face stimuli, or whether researchers were beginning to use spontaneous or dynamic stimuli relatively more frequently, and posed or still images less. Note, percentages may exceed 100 when summed across categories because some articles used more than one type of stimulus.

RESULTS

Our search identified 3374 articles that used faces as stimuli from 2000 to 2020 in our 12 journals. Table 1 shows the vast majority used photographic-quality stimuli, with 90% of articles using such stimuli for the 12 journals combined. This pattern was also consistent across the individual journals: for 11 of the 12 journals, more than 80% of articles used photo stimuli. The only exception was *Computers in Human Behavior*, for which articles were less likely to use photo stimuli (67%) and more likely to use virtual stimuli (49%) than other journals. The use of virtual stimuli ranged from 0 to 13% across the remaining 11 journals, and the use of other non-photo stimuli ranged from 2 to 12% across all journals.

Fig. 3 The top row shows data for all 12 journals combined, across 2000 to 2020. All other rows show data for each journal separately, across 2000 to 2020. **A** Blue panels show the number of articles testing face stimuli. **B** Red panels show the total number of Scopus records for the same period for each journal. **C** Green panels show the percentage of articles testing face stimuli relative to the total number of articles published (i.e., blue/red). Shaded areas show 95% confidence bands for fits. Data illustrated here are provided in Supplement S3. *Note.* Visual observation of the scatterplots indicated a nonlinear fit might be more appropriate than a linear one for some data. For example, the top left (blue) and top right (green) panels show steady increases from 2000 to 2011 and broadly stabilize or decrease thereafter. Statistical analysis confirmed this pattern was better fit by a quadratic than a linear function, and thus we present both types of fit for these plots. ^Indicates that a quadratic fit explained >10% more variance than the linear fit for this data, and thus we present the quadratic fit and associated R^2 . Supplement S3 also reports the linear and quadratic fit R^2 for data in each plot

Are face stimuli being used more or less in research over time?

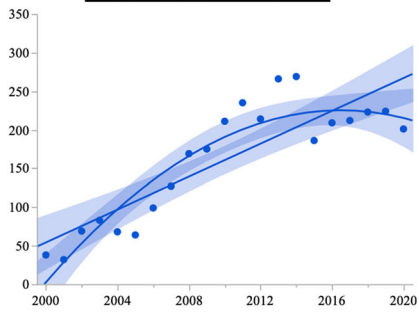
The top row of Fig. 3 illustrates the numbers and percentages of articles that used faces as stimuli across 2000 to 2020 for all 12 journals combined; the remaining rows do likewise for each journal separately. The absolute number of articles that used faces as stimuli increased across 2000 to ~2014 and stabilized thereafter, as indicated by the quadratic fit for the combined journals data, $R^2 = .87$, $p < .001$ (top row of Fig.

Table 1 Percentage and number of articles that used faces as stimuli, and percentage of articles that used different types of face stimuli for each journal

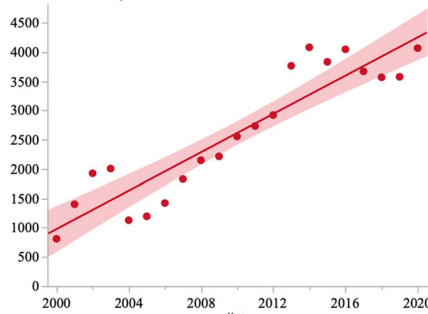
Journal	% of records using face stimuli ^a	No. articles using face stimuli	% of articles using different types of face stimuli ^b		
			Photo ^c	Virtual/avatar	Other non-photo ^d
<i>Cognition</i>	5%	182	86%	10%	8%
<i>Cognition and Emotion (C&E)</i>	18%	372	90%	3%	9%
<i>Computers in Human Behavior (CHB)</i>	1%	43	67%	49%	12%
<i>Emotion</i>	22%	406	89%	7%	8%
<i>Frontiers in Human Neuroscience (FrHN)</i>	4%	285	91%	7%	8%
<i>Journal of Affective Disorders (JAD)</i>	1%	116	97%	0%	3%
<i>Journal of Personality and Social Psychology (JPSP)</i>	3%	100	90%	12%	9%
<i>Journal of Vision (JoV)</i>	4%	275	83%	13%	9%
<i>Neuropsychologia (NPLogia)</i>	10%	649	95%	3%	5%
<i>Perception</i>	14%	379	88%	7%	11%
<i>Psychological Science (Psych Sci)</i>	6%	241	92%	9%	2%
<i>Social Cognitive and Affective Neuroscience (SCAN)</i>	19%	326	92%	7%	2%
All journals	6%	3374	90%	7%	7%

Note. ^a N articles that used faces as stimuli/ N total records returned by Scopus for each journal for 2000–2020. ^b Rows may add up to >100% because some articles used more than one type of face stimuli. ^c Photographic-quality images or videos of real human faces, including digitally manipulated photographs and videos. ^d Includes painted, line drawing, and schematic representations of human faces.

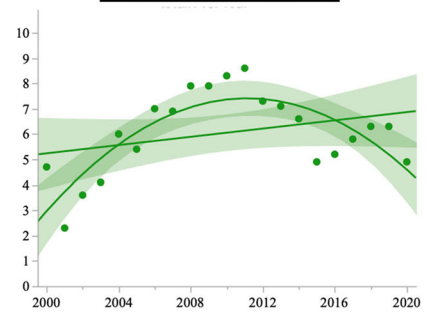
A *N* articles used faces as stimuli
 linear $r/R^2 = .86/.74^{***}$
 Δ quadratic $R^2 = .87^{***}$



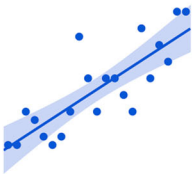
B *N* total records
 linear $r/R^2 = .92/.85^{***}$
 quadratic $R^2 = .85^{***}$



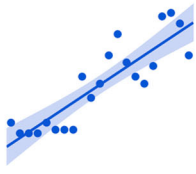
C % total records used faces as stimuli
 linear $r/R^2 = .31/.10$
 Δ quadratic $R^2 = .64^{***}$



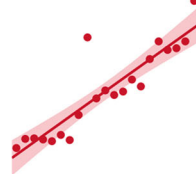
Cognition $R^2 = .67^{***}$



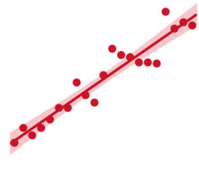
C&E $R^2 = .77^{***}$



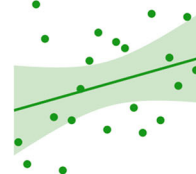
Cognition $R^2 = .81^{***}$



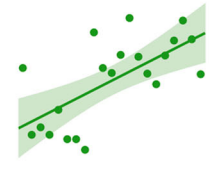
C&E $R^2 = .91^{***}$



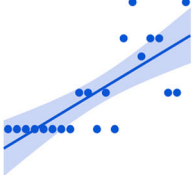
Cognition $R^2 = .10$



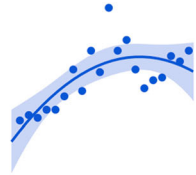
C&E $R^2 = .44^{**}$



CHB $R^2 = .56^{***}$



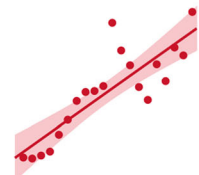
Δ Emotion $R^2 = .58^{***}$



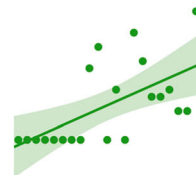
CHB $R^2 = .68^{***}$



Emotion $R^2 = .72^{***}$



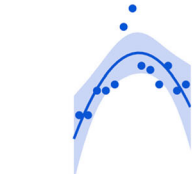
CHB $R^2 = .36^{**}$



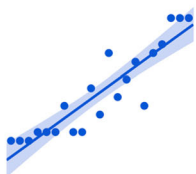
Δ Emotion $R^2 = .29, p = .05$



Δ FrHN $R^2 = .55^*$



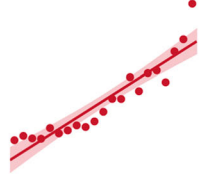
JAD $R^2 = .83^{***}$



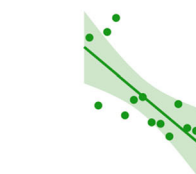
Δ FrHN $R^2 = .72^{**}$



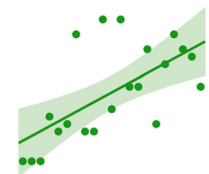
JAD $R^2 = .86^{***}$



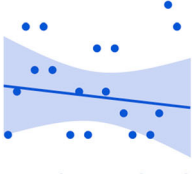
FrHN $R^2 = .52^{**}$



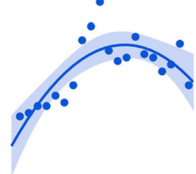
JAD $R^2 = .41^{**}$



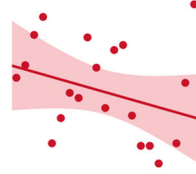
JPSP $R^2 = .01$



Δ JoV $R^2 = .65^{***}$



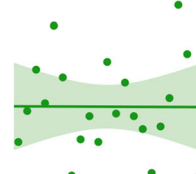
JPSP $R^2 = .10$



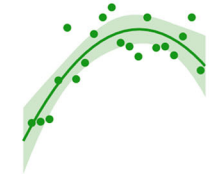
Δ JoV $R^2 = .28, p = .06$



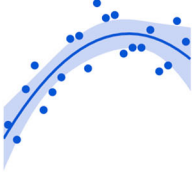
JPSP $R^2 = .00$



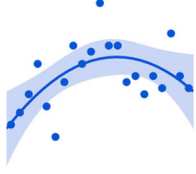
Δ JoV $R^2 = .66^{***}$



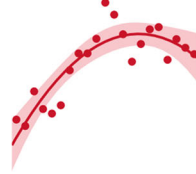
Δ NPlogia $R^2 = .65^{***}$



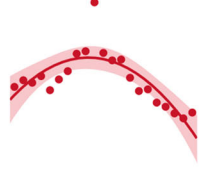
Δ Perception $R^2 = .36^*$



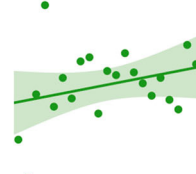
Δ NPlogia $R^2 = .77^{***}$



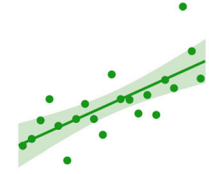
Δ Perception $R^2 = .64^{***}$



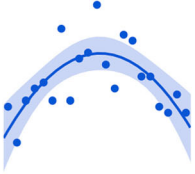
NPlogia $R^2 = .09$



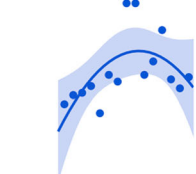
Perception $R^2 = .53^{***}$



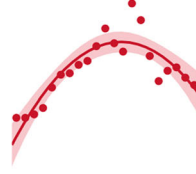
Δ Psych Sci $R^2 = .52^{**}$



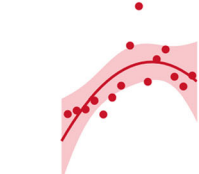
Δ SCAN $R^2 = .38, p = .06$



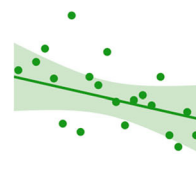
Δ Psych Sci $R^2 = .78^*$



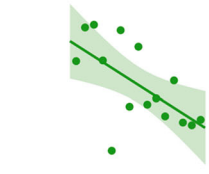
Δ SCAN $R^2 = .49^*$



Psych Sci $R^2 = .11$



SCAN $R^2 = .39^*$



3A). Seven of the 12 individual journals followed this quadratic pattern. Four of the five remaining journals showed significant linear increases that continued from 2000 through to 2020 (*Cognition*, *Cognition and Emotion*, *Computers in Human Behavior*, *Journal of Affective Disorders*). The final journal showed a small but statistically nonsignificant decrease over time (*Journal of Personality and Social Psychology*).

However, Fig. 3B shows these changes occurred within the context of an overall massive increase in the total number of articles the 12 journals published each year, as indicated by significant linear fits for the combined journals data, $r = .92$, $p < .001$ (top row of Fig. 3B), and for 9 of the 12 individual journals, $r_s = .60$ to $.95$, all $p_s < .05$. Several of the journals also showed quadratic trends that suggested their publication rates have stabilized or even decreased in the last few years (most notably for *Neuropsychologia*, *Perception*, and *Psychological Science*; see small red panels in Fig. 3B and Supplement 3 for details).

Thus, to investigate whether there were changes in the number of articles that tested face stimuli relative to overall changes in publication rates, we calculated the percentage of total articles in our 12 journals that tested face stimuli (i.e., Fig. 3: blue/red $\times 100$ = green data). For the combined journals data, we found the percentage of articles that used faces as stimuli fitted a quadratic pattern, increasing from 4.7% in 2000 to peak at 8.9% in 2011, and then decreasing back to 4.9% in 2020, $R^2 = .64$, $p < .001$ (top row of Fig. 3C). For individual journals, the results were mixed. A similar quadratic pattern was evident for only two journals (*Emotion* $R^2 = .29$, $p = .051$; *Journal of Vision*

$R^2 = .66$, $p < .001$). Four of the other journals (*Cognition and Emotion*, *Computers in Human Behavior*, *Journal of Affective Disorders*, *Perception*) showed significant linear increases in the percentage of articles that used faces as stimuli over time, $r_s = .57$ to $.73$, all $p_s < .05$. However, for the two key neuroscience journals (*Frontiers in Human Neuroscience*, *Social Cognitive and Affective Neuroscience*) the percentage of articles that used faces as stimuli decreased linearly over time, $r_s = -.63$ to $-.72$, both $p_s < .05$.

How much of the literature has used standardized face stimuli relative to ambient ones?

We identified 3040 articles that used photographic-quality stimuli (Supplement S4), the vast majority of which—2415 articles (79%)—used standardized stimuli. In contrast, 357 (12%) reported using ambient stimuli. Four hundred and thirty-eight articles (14%) did not provide enough information about one or more of their stimulus sets to judge whether they were standardized or ambient. Figure 4 converts these numbers into the percentage of articles for each year that used each type of photographic-quality stimulus. While the percentage of articles that did not provide sufficient information to judge whether stimuli were standardized or ambient has significantly decreased over the past two decades, $r = -.87$, $p < .001$, there has been a concurrent significant increase in the percentage of articles that report using standardized stimuli, $r = .66$, $p = .001$. The percentage of articles using ambient stimuli has remained stable, averaging around 12%, $r = -.05$, $p = .833$.

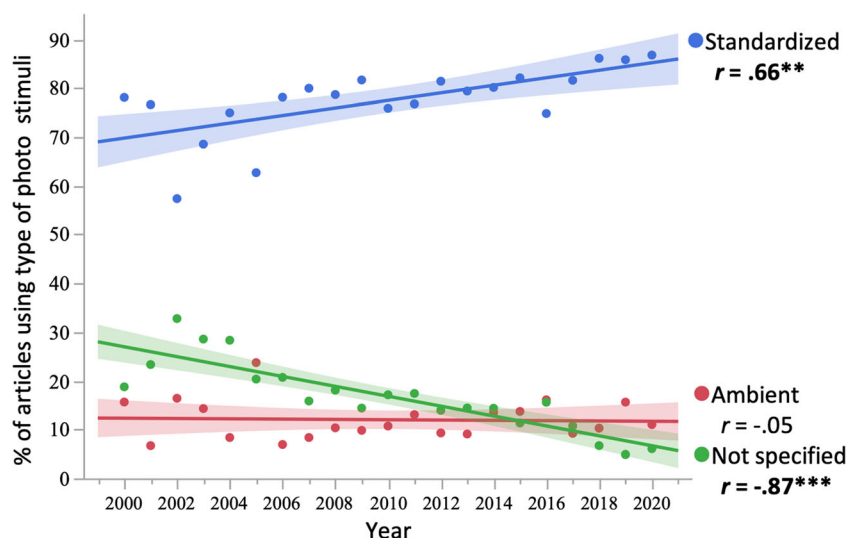


Fig. 4 Percentage of articles using standardized, ambient, and unspecified types of photo stimuli by year

How much of face expression/emotion work has used posed expressions?

We identified 1569 articles that addressed questions about expression/emotion using photographic-quality stimuli (Supplement S5). Of these, the vast majority reported using posed expressions ($n = 1398$, 89%). Although only a small percentage of articles reported using spontaneous expressions ($n = 119$, 8%), this percentage has increased significantly over time, $r = .74$, $p < .001$. One hundred and seventy-eight articles (11%) did not provide enough information about one or more of their stimulus sets to judge whether they were posed or spontaneous.

Figures 5A and 5B illustrate the number and percentage of the 1569 expression/emotion articles that used the four most popular expression sets, which were, as anticipated, Ekman and colleagues' (used in 30% of articles), followed by the KDEF (23%), RaFD (21%), and NimStim (7%). Considering these four sets in combination, the absolute number of articles that used one or more of them increased significantly across 2000 to 2020, $r = .88$, $p < .001$ (orange line in Fig. 5A). However, there was no change in how much they were used relative to other types of photo expression stimuli: the percentage of articles that used one or more of these four sets remained stable across 2000 to 2020, $r = -.07$, $p = .845$

(orange line in Fig. 5B), hovering around 70%. There were changes in how much the four sets were used relative to one another though. The percentage of articles using Ekman and colleagues' stimuli decreased over time, $r = -.95$, $p < .001$ (purple line in Fig. 5B), while the percentage using the three newer sets increased, $r = .91$, $p < .001$ (data for KDEF, NimStim and RaFD combined; see Fig. 5 for r for each set).

Still versus dynamic presentation of facial expressions

Of the 1569 articles that addressed questions about expression/emotion using photographic-quality stimuli, the vast majority used static expressions (1430, 91%). Only 198 articles (13%) used dynamic expressions, of which a quarter ($n = 51$) used dynamic morphs (Supplement S7). That is, stitched-together sequences of morphed images (e.g., showing from neutral to happy in 20% morph increments, with images in the sequence shown for 200 ms each; Putman et al., 2006), so that the face appeared to be moving from neutral to an expression, or from one expression to another. Although the number of articles that used dynamic expressions increased significantly across 2000 to 2020, $r = .86$, $p < .001$, the percentage relative to the total number of expression/emotion articles each year did not, $r = .18$, $p = .427$.

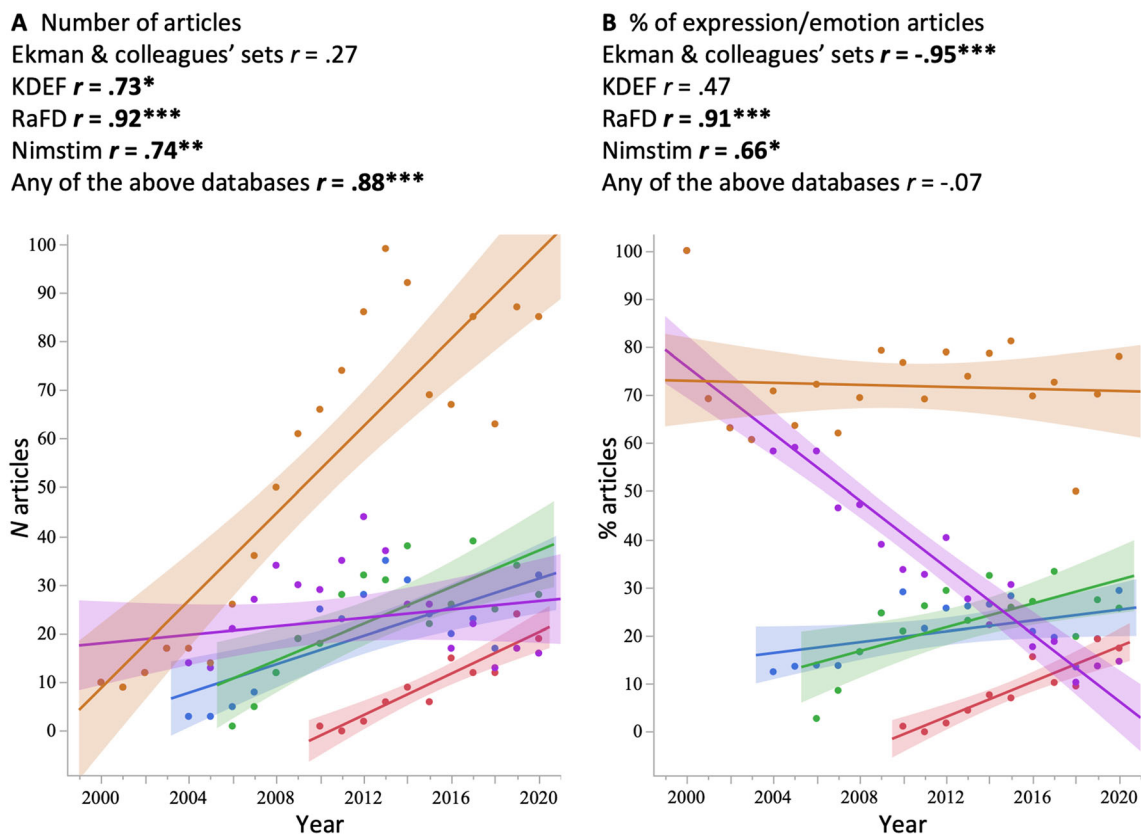


Fig. 5 **A** Number and **B** percentage of articles that addressed questions about expression/emotion using photographic-quality stimuli, for the four most popular databases. Shaded areas show 95% confidence bands for linear fits. The data illustrated here are provided in Supplement S6

Virtual face stimuli

In total, 247 articles used virtual face stimuli (Supplement S8). This number equates to 7.3% of the total articles included in our survey, indicating virtual faces are being used in a small but nontrivial amount of research. However, we found no articles had used GAN faces. Regarding the extent to which people are using virtual faces as stand-ins for images showing real humans (i.e., to supposedly understand human responses to human faces), this was the case for the vast majority of articles ($n = 191$, 77%).

Regarding changes over time, we found the use of virtual faces has increased. In the early 2000s, there were very few articles published; most appeared in the last decade. Importantly, correlational analyses established the increased use of virtual faces across 2000 to 2020 was significant, for the percentage of articles using virtual faces relative to the total number of articles included in our study each year, $r = .82$, $p < .001$, as well as for the absolute number of articles, $r = .87$, $p < .001$.

DISCUSSION

The findings of the present survey highlight that psychological knowledge about face processing is lacking input from stimuli that capture the natural and broad repertoire of real-life facial behavior. Most studies have used photographic-quality stimuli, but the overwhelming majority have been standardized to eliminate the natural variation in pose, illumination, etc., of real-life faces. After taking into account the massive increases in publishing rates since the turn of the century, we found the use of ambient faces has remained stable over time at around 12%. The use of dynamic (moving) expressions has also remained stable, at around 13%. The use of spontaneous expressions has increased, but posed expressions continue to dominate research. The four most popular facial expression databases—all of which comprise static images of standardized, posed expressions—have consistently been used in around 70% of expression articles across the past two decades, although the relative use of Ekman and colleagues' sets has declined in favor of newer ones (KDEF, NimStim, RaFD). Some possible reasons for the shift towards the newer sets include that they are colored rather than grayscale¹, show faces from multiple viewpoints (KDEF, RaFD), with different eye-gaze directions (RaFD), or systematically varied how expressions were posed (i.e., the NimStim set includes open- and closed-mouthed versions of some expressions). An emerging trend is the increasing use of virtual face stimuli, often as proxies for human faces in research. A final point of note is that, while some journals have published

relatively more articles that use faces as stimuli over time (i.e., *Cognition*, *Cognition and Emotion*, *Computers in Human Behavior*, *Journal of Affective Disorders*), the use of face stimuli in the two neuroscience journals we surveyed has declined.

Why are ambient and spontaneous stimuli not used more?

There is much to be gained by balancing the use of traditional face stimuli with ones that better capture the full range of how we see faces in real life. Why, then, are such stimuli not more popular? One reason is that, until recently, it has been difficult to obtain and present naturalistic stimuli (e.g., due to insufficient computing power). However, social and technological advances have opened up access to stimuli that capture our real-life diet of faces. For example, researchers are starting to make use of the material available on YouTube (e.g., Wenzler et al., 2016) and mobile eye-trackers to study natural interactions (Foulsham, 2015). Good dynamic expression databases, including spontaneous ones, are also increasingly available (Krumhuber et al., 2017).

A second reason for preferring standardized over ambient stimuli is psychology's focus on the experimental method. The experimental method advocates using highly controlled stimuli so that the parameters of interest can be cleanly manipulated to pinpoint causal relationships. Within this context, ambient stimuli are often not appropriate. For example, early tests of face recognition ability failed because they included ambient cues (e.g., hair) that could be used to "cheat" the test so that even people with very poor face recognition ability (i.e., prosopagnosics) performed well (Duchaine & Nakayama, 2006). Also, some scientific paradigms have been developed to work specifically with standardized stimuli. Adaptation paradigms, used to understand the neural coding of faces, often use morphed stimuli that must be generated from standardized images of the same person (e.g., showing different expressions; Skinner & Benton, 2010; Palermo et al., 2018). Other paradigms have been developed to work with static images, and may be difficult to adapt to dynamic stimuli (but see study 4 of Grafton et al., 2021 for an example of how this can be done and improves reliability). Altogether, these examples illustrate that there are times when using standardized and posed stimuli is beneficial, and even necessary.

However, some of the supposedly "extraneous" variation that is eliminated in standardized and posed stimuli may be an inherent part of the phenomena psychology seeks to understand. For instance, spontaneous expressions are much more variable and may include different facial actions than those depicted in standardized sets (Krumhuber et al., 2021; Matsumoto et al., 2009; Naab & Russell, 2007; Namba, Kagamihara, et al., 2017a; Namba, Makihara, et al., 2017b). In such cases, an approach that seeks convergent evidence

¹ Ekman and colleagues' stimuli are grayscale, not colored.

from controlled stimuli *and* “in the wild” stimuli *and* natural interactions may be more useful. Kingstone et al. (2008; also see Kingstone, 2020) coined this combined approach *cognitive ethology*. The strength of cognitive ethology is that observation of phenomena in the natural world is explicitly used to inform experimental testing and vice versa. There are also strategies that can be used to mitigate concerns about experimental confounds from ambient stimuli. For example, Rossion et al. (2015) advocate an EEG approach that aims to isolate what is common across images (the presence of a face) by using large numbers of ambient images, so that the irrelevant variation in low-level cues is random and does not have any systematic influence on the effect of interest.

Implications of the stimulus imbalance

The low use of ambient and spontaneous expression stimuli has potentially far-reaching implications for psychological theories of face processing, including theories of atypical face processing (e.g., as has already been found for schizophrenia; Davis & Gibson, 2000; and psychopathic traits, Dawel, Wright, et al., 2019b). For example, the potential disparity between lab and real-world processes might explain an apparent paradox in autism spectrum disorder (ASD). The clinical profile of people with ASD is characterized by impairments in social functioning (American Psychiatric Association, 2013). Yet lab-based studies frequently find people with ASD do not have any difficulties labeling facial expressions (Harms et al., 2010). Harms et al. (2010) suggest that people with ASD might use compensatory strategies to perform lab-based emotion labeling tasks. We believe that using “in the wild” stimuli that present spontaneous and dynamic facial behavior under ambient conditions is critical for understanding social cognition problems in ASD and other clinical disorders.

Another example concerns how psychological theories derived from posed expressions have been applied in computer science to supposedly “read” people’s emotions (e.g., <https://www.affectiva.com/>). This software is programmed to identify the facial movements set out in the Directed Facial Action Task (Ekman, 2007) used to pose Ekman and colleagues’ stimuli (Ekman & Friesen, 1976; Matsumoto & Ekman, 1988; Young et al., 2002) and the RaFD (Langner et al., 2010). It is unsurprising then that this software can be easily fooled into identifying someone as feeling an emotion when they are not (see: <https://www.theguardian.com/technology/2021/apr/04/online-games-ai-emotion-recognition-emojify>) and performs worse for spontaneous than posed expressions (Krumhuber et al., 2019; Krumhuber et al., 2021). Using more spontaneous expressions to inform this technology may improve its usefulness.

The rise of virtual faces

Virtual faces were rare in psychological research in the early 2000s, but have become more common over the last decade. One potential problem is that experimental work is mostly using virtual face stimuli as proxies for human faces, to address questions about human responses to human faces. The appeal of this approach is that virtual faces are extremely well standardized, and enable experimenters to cleanly manipulate whatever parameters are of interest. However, the evidence so far suggests that people respond differently to virtual and human faces (e.g., Balas & Pacella, 2015; Crookes et al., 2015; Katsyri et al., 2019; Schindler et al., 2017). A related concern is that software can be used to manipulate virtual faces beyond the bounds of real facial behavior (e.g., combining AUs or using timing parameters that do not occur in real life), risking findings that cannot be replicated for real human facial behavior. However, the extent to which findings for human faces are replicated with virtual ones may depend on the type of virtual face being used. All virtual faces included in our survey were virtual-looking (e.g., created using FaceGen or Poser software). No studies used the newer, more realistic GAN faces, which are only just starting to appear in the literature. GAN faces might be appropriate replacements for human faces in research, but we currently lack an evidence base to support this. In addition, investigating how humans perceive and respond to virtual faces is of interest in its own right, given their increasing role in our daily lives (e.g., Andrade et al., 2010; Dellazizzo et al., 2018; Guise et al., 2012).

Limitations

An important limitation of the present survey is that it only describes patterns in the use of face stimuli, and does not identify the factors that are causing these patterns. We have considered what some of these factors might be in our discussion, but our reasoning is based on logic and common sense. A second limitation of the present survey is that a large number of studies did not provide sufficient detail about stimuli in their paper for coding. However, the use of standardized, posed, and non-moving stimuli were the overwhelming majority so that even had the articles that did not provide sufficient detail used ambient, spontaneous, and dynamic stimuli, our key findings would still stand. A third limitation is that we only included articles published since the turn of the century. It is likely that trends before 2000 are very different. For example, very few expression databases were available before 2000, and virtual stimuli had not yet emerged in the literature.

Conclusion

Overall, the results of the present survey highlight that there has been a significant imbalance in favor of standardized face

stimuli over naturalistic ones so that the full repertoire of faces and facial behavior people see in everyday life has been underrepresented. One reason for this imbalance is that ambient and spontaneously expressive stimuli have traditionally been challenging to access and validate. However, new opportunities are opening up. Another is that experimental control is reduced for ambient relative to standardized stimuli, making it more difficult to draw causal inferences. However, multiple studies show that using ambient as well as standardized stimuli, and spontaneous as well as posed expressions, will help build a fuller and more accurate understanding of human face processing. Ideally, lab research should combine correlational studies of naturalistic faces with rigorous experimental work using standardized ones. Both types of stimuli are useful, but for different reasons. It would benefit the literature to be more explicit about the rationale for using a given stimulus type to answer whatever research question is at hand. More broadly, adopting a cognitive ethological approach that extends work on human face processing to natural settings should improve theoretical understanding.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13428-021-01705-3>.

Acknowledgments We thank Ziyun Wu and William Whitecross for their help with coding the articles.

References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. American Psychiatric Association: Arlington, USA.
- 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.
- Aviezer, H., Messinger, D. S., Zangvil, S., Mattson, W. I., Gangi, D. N., & Todorov, A. (2015). Thrill of victory or agony of defeat? Perceivers fail to utilize information in facial movements. *Emotion*, *15*(6), 791–797.
- Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*(6111), 1225–1229.
- Balas, B., & Pacella, J. (2015). Artificial faces are harder to remember. *Computers in Human Behavior*, *52*, 331–337.
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, *20*(1), 1–68.
- Blair, R. J. R. (2006). The emergence of psychopathy: Implications for the neuropsychological approach to developmental disorders. *Cognition*, *101*(2), 414–442.
- Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1467–1485.
- 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.
- Crookes, K., Ewing, L., Gildenhuis, J. D., Kloth, N., Hayward, W. G., Oxner, M., . . . Rhodes, G. (2015). How well do computer-generated faces tap face expertise? *PLoS One*, *10*(11), e0141353.
- Davis, P. J., & Gibson, M. G. (2000). Recognition of posed and genuine facial expressions of emotion in paranoid and nonparanoid schizophrenia. *Journal of Abnormal Psychology*, *109*(3), 445–450.
- Dawel, A., Dumbleton, R., O’Kearney, R., Wright, L., & McKone, E. (2019a). Reduced willingness to approach genuine smilers in social anxiety explained by potential for social evaluation, not misperception of smile authenticity. *Cognition and Emotion*, *33*(7), 1342–1355.
- Dawel, A., Palermo, R., O’Kearney, R., & McKone, E. (2015). Children can discriminate the authenticity of happy but not sad or fearful facial expressions, and use an immature intensity-only strategy. *Frontiers in Psychology*, *6*, 462.
- Dawel, A., Wright, L., Dumbleton, R., & McKone, E. (2019b). All tears are crocodile tears: Impaired perception of emotion authenticity in psychopathic traits. *Personality Disorders: Theory, Research, and Treatment*, *10*(2), 185–197.
- Dawel, A., Wright, L., Irons, J., Dumbleton, R., Palermo, R., O’Kearney, R., & McKone, E. (2017). Perceived emotion genuineness: Normative ratings for popular facial expression stimuli and the development of perceived-as-genuine and perceived-as-fake sets. *Behavior Research Methods*, *49*(4), 1539–1562.
- 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.
- Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585.
- Ekman, P. (1980). Asymmetry in facial expression. *Science*, *209*, 833–834.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200.
- Ekman, P. (2007). The directed facial action task: Emotional responses without appraisal. In Coan, J.A. & Allen, J.J.B. (Eds), *Handbook of emotional elicitation and assessment* (pp. 47–53). : Oxford University Press.
- Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. : Consulting Psychologists Press.
- Foulsham, T. (2015). Eye movements and their functions in everyday tasks. *Eye*, *29*(2), 196–199.
- Grafton, B., Teng, S., & MacLeod, C. (2021). Two probes and better than one: Development of a psychometrically reliable variant of the attentional probe task. *Behavior Research and Therapy*, *138*, 103805.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). Generative Adversarial Networks. *Proceedings of the International Conference on Neural Information Processing Systems*, 2672–2680.
- Guise, V., Chambers, M., Conradi, E., Kavia, S., & Valimaki, 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.
- Gunnery, S. D., & Ruben, M. A. (2016). Perceptions of Duchenne and non-Duchenne smiles: A meta-analysis. *Cognition and Emotion*, *30*(3), 501–515.
- Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*(3), 290–322.
- Heerey, E. A. (2014). Learning from social rewards predicts individual differences in self-reported social ability. *Journal of Experimental Psychology: General*, *143*(1), 332–339.

- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323.
- Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4401–4410.
- Katsyri, J., de Gelder, B., & de Borst, A. W. (2019). Amygdala responds to direct gaze in real but not in computer-generated faces. *Neuroimage*, *204*, 116216.
- Kingstone, A. (2020). Everyday human cognition and behavior. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale* *74*(4), 267–274.
- Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340.
- Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2010). Facial emotion perception in schizophrenia: A meta-analytic review. *Schizophrenia Bulletin*, *36*(5), 1009–1019.
- Krumhuber, E. G., Küster, D., Namba, S., Shah, D., & Calvo, M. G. (2019). Emotion recognition from posed and spontaneous dynamic expressions: Human observers versus machine analysis. *Emotion*
- Krumhuber, E. G., Küster, D., Namba, S., & Skora, L. (2021). Human and machine validation of 14 databases of dynamic facial expressions. *Behavior Research Methods*, *53*(2), 686–701.
- Krumhuber, E. G., Skora, L., Küster, D., & Fou, L. (2017). A review of dynamic datasets for facial expression research. *Emotion Review*, *9*(3), 280–292.
- 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.
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388.
- LaRusso, L. (1978). Sensitivity of paranoid patients to nonverbal cues. *Journal of Abnormal Psychology*, *87*(5), 463–471.
- Levenson, R., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*(4), 363–384.
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). About KDEF. Retrieved from <http://kdef.se/home/aboutKDEF.html>
- Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian Facial Expressions of Emotion (JACFEE) [Photos]*. San Francisco, CA: Intercultural and Emotion Research Laboratory, Department of Psychology, San Francisco State University.
- Matsumoto, D., Olide, A., Schug, J., Willingham, B., & Callan, M. (2009). Cross-cultural judgments of spontaneous facial expressions of emotion. *Journal of Nonverbal Behavior*, *33*(4), 213–238.
- 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*
- Naab, P. J., & Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion*, *7*(4), 736.
- Namba, S., Kagamihara, T., Miyatani, M., & Nakao, T. (2017a). Spontaneous facial expressions reveal new action units for the sad experiences. *Journal of Nonverbal Behavior*, *41*(3), 203–220.
- Namba, S., Makihara, S., Kabir, R. S., Miyatani, M., & Nakao, T. (2017b). Spontaneous facial expressions are different from posed facial expressions: Morphological properties and dynamic sequences. *Current Psychology*, *36*(3), 593–605.
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092.
- Palermo, R., Jeffery, L., Lewandowsky, J., Fiorentini, C., Irons, J. L., Dawel, A., ... & Rhodes, G. (2018). Adaptive face coding contributes to individual differences in facial expression recognition independently of affective factors. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(4), 503.
- Putman, P., Hermans, E., & Van Honk, J. (2006). Anxiety meets fear in perception of dynamic expressive gaze. *Emotion*, *6*(1), 94.
- Ritchie, K. L., Smith, F. G., Jenkins, R., Bindemann, M., White, D., & Burton, A. M. (2015). Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-ID paradox. *Cognition*, *141*, 161–169.
- Rossion, B., Torfs, K., Jacques, C., & Liu-Shuang, J. (2015). Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *Journal of Vision*, *15*(1), 18–18.
- Russell, J. A., & Fernández-Dols, J. M. (Eds.). (1997). *The psychology of facial expression*. : Cambridge University Press.
- 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), 45003–45003.
- Skinner, A. L., & Benton, C. P. (2010). Anti-expression aftereffects reveal prototype-referenced coding of facial expressions. *Psychological Science*, *21*(9), 1248–1253.
- Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118.
- Shore, D. M., & Heerey, E. A. (2011). The value of genuine and polite smiles. *Emotion*, *11*(1), 169–174.
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460.
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249.
- Wenzler, S., Levine, S., van Dick, R., Oertel-Knöchel, V., & Aviezer, H. (2016). Beyond pleasure and pain: Facial expression ambiguity in adults and children during intense situations. *Emotion*, *16*(6), 807–814.
- Young, A., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion—Stimuli and tests (FEEST)*. : Thames Valley Test Company.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.