

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/330296845>

# Words Are a Context for Mental Inference

Article in *Emotion* · January 2019

DOI: 10.1037/emo0000510

CITATIONS

0

READS

384

3 authors:



Nicole Betz

Northeastern University

8 PUBLICATIONS 25 CITATIONS

SEE PROFILE



Katie Hoemann

Northeastern University

10 PUBLICATIONS 25 CITATIONS

SEE PROFILE



Lisa Feldman Barrett

Northeastern University

309 PUBLICATIONS 24,053 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Super aging: Behavioral characteristics and neural substrates of "successful" cognitive-affective aging [View project](#)



Society for Affective Science Students [View project](#)

# Emotion

## Words Are a Context for Mental Inference

Nicole Betz, Katie Hoemann, and Lisa Feldman Barrett

Online First Publication, January 10, 2019. <http://dx.doi.org/10.1037/emo0000510>

### CITATION

Betz, N., Hoemann, K., & Barrett, L. F. (2019, January 10). Words Are a Context for Mental Inference. *Emotion*. Advance online publication. <http://dx.doi.org/10.1037/emo0000510>

# Words Are a Context for Mental Inference

Nicole Betz and Katie Hoemann  
Northeastern University

Lisa Feldman Barrett  
Northeastern University and Massachusetts General Hospital,  
Boston, Massachusetts

Accumulating evidence indicates that context has an important impact on inferring emotion in facial configurations. In this paper, we report on three studies examining whether words referring to mental states contribute to mental inference in images from the Reading the Mind in the Eyes Test (Study 1), Baron-Cohen et al. (2001) in static emoji (Study 2), and in animated emoji (Study 3). Across all three studies, we predicted and found that perceivers were more likely to infer mental states when relevant words were embedded in the experimental context (i.e., in a forced-choice task) versus when those words were absent (i.e., in a free-labeling task). We discuss the implications of these findings for the widespread conclusion that faces or parts of faces “display” emotions or other mental states, as well as for psychology’s continued reliance on forced-choice methods.

**Keywords:** categorical perception, emotion perception, research methodology

**Supplemental materials:** <http://dx.doi.org/10.1037/emo0000510.supp>

People cry in happiness at weddings, in sadness at funerals, and in frustration at computer crashes, so how does a perceiver know whether a cry signals happiness, sadness, or frustration? The answer is context (Barrett, Adolphs, Marsella, Martinez, & Pollak, 2018; Barrett, Mesquita, & Gendron, 2011). In many psychology experiments, words and their associated conceptual knowledge provide an unintended context for establishing above-chance levels of agreement that serve as estimates of ‘accuracy’ (for reviews, see Barrett, 2011; Barrett, Lindquist, & Gendron, 2007; Gendron, Mesquita, & Barrett, 2013; Russell, 1994). For example, most studies of emotion perception use a forced-choice method to assess people’s ability to assign an emotional meaning to a facial configuration. On a given trial, perceivers are typically shown a single cue—often a photograph of an isolated posed facial configuration, such as a wide-eyed gasping face, a scowling face, or a smiling face—along with a small selection of emotion words. Perceivers are then tasked with labeling the face, and agreement with the experimenter’s hypothesis is computed as an index of emotion recognition ‘accuracy’ (e.g., Ekman & Friesen, 1986; Keltner, 1996; Tracy & Robins, 2004, Experiment 1). Perceivers assign these posed photos into emotion categories at above-chance levels

of agreement (for a meta-analysis, see Elfenbein & Ambady, 2002; for recent reviews, see Keltner, Sauter, Tracy, & Cowen, in press; Shariff & Tracy, 2011).

However, agreement is substantially reduced, often to chance levels, by removing the words or otherwise limiting access to emotion concept knowledge during the task. This can be achieved by asking participants to freely label the face cues (e.g., Boucher & Carlson, 1980; Russell, 1997; for reviews, see Barrett et al., 2018; Gendron, Crivelli, & Barrett, 2018), by semantically satiating emotion concepts and then observing impaired repetition priming (Gendron, Lindquist, Barsalou, & Barrett, 2012) and reduced perceptual matching (Lindquist, Barrett, Bliss-Moreau, & Russell, 2006), or by testing individuals with disorders that impair conceptual processing, such as semantic dementia (Lindquist, Gendron, Barrett, & Dickerson, 2014) or semantic aphasia (Roberson, Davidoff, & Braisby, 1999). When the conceptual context is substantially reduced, physical actions seem more ambiguous and a variety of mental states are inferred (e.g., Aviezer et al., 2008; Aviezer, Trope, & Todorov, 2012; de Gelder, 2006; Hassin, Aviezer, & Bentin, 2013; Wieser & Brosch, 2012). When forced-choice tasks are used to study emotion perception in remote cultural samples, agreement rates are consistently above chance, and are interpreted as evidence that certain facial configurations are universally recognized as emotional expressions. By contrast, asking participants to freely label the same photos consistently fails to replicate these above-chance levels of agreement, and therefore calls this interpretation into question (Gendron et al., 2018; for specific examples, see Crivelli, Jarillo, Russell, & Fernandez-Dols, 2016; Gendron, Roberson, van der Vyver, & Barrett, 2014a, 2014b).

Despite the evidence that words and other conceptual content may influence emotion perception, forced-choice tasks are in widespread use in psychology (e.g., Atkinson, Dittrich, Gemmell, & Young, 2004; Cordaro, Keltner, Tshering, Wangchuk, & Flynn, 2016; Tracy, Robins, & Schriber, 2009), neuroscience (for a re-

---

Nicole Betz and Katie Hoemann, Department of Psychology, Northeastern University; Lisa Feldman Barrett, Department of Psychology, Northeastern University, and Psychiatric Neuroimaging Division, Department of Psychiatry, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, Massachusetts.

This research was supported by the National Institutes of Health Director’s Pioneer Award DP1OD003312 to Lisa Feldman Barrett.

Correspondence concerning this article should be addressed to Nicole Betz or Lisa Feldman Barrett, Department of Psychology, Northeastern University, 125 Nightingale Hall, 360 Huntington Avenue, Boston, MA 02115. E-mail: betz.n@husky.neu.edu or l.barrett@neu.edu

view, see Brooks et al., 2017) and related scientific fields (e.g., M. Dyck et al., 2008; Tinwell, Grimshaw, Nabi, & Williams, 2011). For example, in an important meta-analysis of emotion perception published in 2002, 162 of 168 effect sizes came from studies using forced-choice tasks (Elfenbein & Ambady, 2002, Table 1). Scientist often interpret their findings without acknowledging their potential impact (for a discussion, see Barrett & Gendron, 2016). In this paper, we build on these prior findings with three experiments designed to test the impact of mental state words for influencing mental inferences during tasks of social perception. We discuss the theoretical implications of these findings, as well as the implications of continuing to rely on forced-choice methods for replicable, generalizable results in psychological science.

### The Impact of Words in Shaping Mental Inferences

There is growing evidence from a broad array of studies that words encourage participants to assign mental meaning to configurations of muscle movements differently than they would if the words were absent from the experimental task. For example, numerous studies now show that emotion words influence how facial configurations are predicted, encoded, and remembered as emotional expressions (e.g., Chanes, Wormwood, Betz, & Barrett, 2018; Doyle & Lindquist, 2018; Fugate, Gendron, Nakashima, & Barrett, 2018; Fugate, Gouzoules, & Barrett, 2010; Halberstadt & Niedenthal, 2001; Nook, Lindquist, & Zaki, 2015). This is in part because facial configurations lack both consistency and specificity for specific emotion categories: many different facial configurations can express the same emotion (low consistency), and the same facial configuration can be used to express different emotions in different situations (low specificity; for a review, see Barrett et al., 2018). Because each emotion category contains instances that vary in their physiological basis (e.g., Siegel et al., 2018), their bodily expressions (e.g., Kleinsmith & Bianchi-Berthouze, 2013), their affective feelings (e.g., Wilson-Mendenhall, Barrett, & Barsalou, 2015), and even their neural basis (e.g., Clark-Polner, Wager, Satpute, & Barrett, 2016; Raz et al., 2016), perceivers rely on context to make meaning out of what are otherwise ambiguous cues. Words are an especially potent context for shaping mental inferences because they are a special type of sensory input that is inextricably linked to concepts and categories (Gelman & Roberts, 2017; Lupyan & Clark, 2015). Simply perceiving a word involves remembering related concept knowledge (Binder, Desai, Graves, & Conant, 2009). Elsewhere, we have hypothesized that conceptual knowledge is a context that categorizes incoming sensory inputs and makes them meaningful, thereby influencing how facial configurations and other sensory signals are understood and acted upon (Barrett, 2006, 2017a, 2017b; Lupyan & Clark, 2015).

Behavioral evidence from developmental psychology further supports the hypothesis that words, such as those used in forced-choice tasks, are psychologically potent. Children who do not understand the meaning of emotion words beyond their affective content also do not infer specific emotional meaning in facial configurations; they only infer affect (Widen, 2013). Furthermore, congenitally deaf children who are born to nonsigning, hearing parents have limited opportunities to learn the meaning of mental state words and are, correspondingly, delayed in making mental inferences about physical cues when compared with hearing chil-

dren or children who are raised by parents who are fluent in sign language (for a review of evidence, see Sidera, Amadó, & Martínez, 2017). Scientists hypothesize that these deaf children's difficulty inferring mental states is primarily attributable to a delay in language learning (e.g., M. J. Dyck & Denver, 2003; Schick, De Villiers, De Villiers, & Hoffmeister, 2007; Spencer & Marschark, 2010).

The impact of emotion words and associated conceptual content on mental inferences is also found in recent evidence that forced-choice tasks involving emotion words may facilitate mental inferences of emotion that would otherwise not occur (for a discussion, see Hoemann et al., in press). For example, participants label scowling faces as "determined" or "puzzled," wide-eyed faces as "hopeful" and gasping faces as "pained" when they are provided with stories about those emotions rather than with stories of *anger*, *surprise*, and *fear* (Carroll & Russell, 1996, Study 2). And when perceivers are given the option to infer social motives in posed portrayals or spontaneous facial expressions of emotion, they do so, often assigning social motives to those faces more consistently than they assign emotion labels (Crivelli, Russell, Jarillo, & Fernández-Dols, 2016, 2017). When participants are free to infer any cause for facial configurations, they often do not perceive an emotion or mental state more generally (for a discussion, see Nelson & Russell, 2013). Participants who are asked to freely label posed faces often perform action identification (Vallacher & Wegner, 1987) and provide a label for behaviors (e.g., 'looking,' see Gendron et al., 2014b) or for aspects of the situation (e.g., 'hurt themselves,' see Gendron et al., 2014b; Widen & Russell, 2013). These other responses are not possible in a forced-choice task yet are critical for gaining a broader understanding of mental inference in physical movements or features.

### The Present Studies

Taken together, the findings and theoretical context illustrate the ongoing need to examine when and how words shape emotion perception and mental inference more broadly. In three studies, we tested the robustness and generalizability of words' influence on mental inference by comparing participants' responses in forced-choice versus free-label tasks using stimuli beyond the traditional, posed face photos used in studies of emotion perception. In Study 1, we examined whether words facilitate mental inference when looking only at another person's eyes using items from the Reading the Mind in the Eyes Test (RMET; Baron-Cohen & Cross, 1992; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Baron-Cohen, Wheelwright, & Jolliffe, 1997). In Studies 2 and 3, we examined whether words influence emotion inferences in caricatures (emoji) that were designed to clearly convey emotions in virtual communication (Zolli, 2015). In all three experiments, we predicted that participants who completed the forced-choice tasks would have higher levels of agreement with the experimenters' expectations (i.e., higher "accuracy") compared with those who completed the free-label tasks.

These three experiments contribute to the literature on mental inference in several important ways. First, they offer the opportunity to evaluate the robustness and generalizability of prior findings showing that words are potent sources of context for mental inference: they test the hypothesis that conceptual clarification confers some benefit when inferring mental meaning in the pho-

tographs of eyes used in psychological experiments and clinical assessment (Study 1) as well as in the caricatured drawings of expressive faces used in electronic communication (Study 2 and 3). Second, our findings suggest the robust power of words to shape perceptions of the social world, and offer a solid observational basis for future studies to test the processes that cause word-guided mental inference, in which context—including experimental context—plays a psychologically active role. These considerations are particularly important in an era when scientists are concerned about the replicability and generalizability of their findings.

### Study 1: Words Influence Performance on the Reading the Mind in the Eyes Test

In Study 1, we examined whether words provide a context for mental inference in the Reading the Mind in the Eyes Test (RMET; Baron-Cohen et al., 2001). This widely used assessment tool features cropped, black-and-white photographs of human eyes presented alongside four mental state words in a forced-choice design; perceivers are tasked with choosing the best word to describe each pair of eyes. The validity of the RMET rests on the assumptions that (a) the eyes display unique information about internal mental states (Baron-Cohen, Wheelwright, et al., 1997) and (b) the ability to read thoughts and feelings in the eyes is innately prepared and universal (Baron-Cohen, Campbell, Karmiloff-Smith, Grant, & Walker, 1995). Poor performance on the RMET is routinely interpreted as indicating a deficit in social cognition because it discriminates adults with high-functioning autism and neurotypical controls (Baron-Cohen et al., 1997, 2001). The RMET has also been used as a measure of emotion perception (e.g., Richell et al., 2003; Tonks, Williams, Frampton, Yates, & Slater, 2007) and empathy (e.g., Rodrigues, Saslow, Garcia, John, & Keltner, 2009).

Few studies have examined the role of words in the RMET. Cassels and Birch (2014) investigated the effect of presence of words on performance on the children's version of the RMET (Baron-Cohen et al., 2001) and found that children performed better on the forced-choice version of the task (49% agreement) than on the free-label version (29% agreement). Nonetheless, the free-label version more reliably discriminated between typically developing children and those with learning disorders (Cassels & Birch, 2014). In comparison with the forced-choice version of the RMET, performance on the free-label version was not significantly correlated with verbal ability (as measured by the Verbal Comprehension test of the Woodcock-Johnson III; Woodcock, McGrew, Mather, & Schrank, 2001), and predicted group differences (i.e., typically developing or learning-disordered) even when controlling for both age and verbal ability.

In Study 1, we compared performance on forced-choice versus free-label versions of the RMET in an adult sample. We hypothesized that the mental state words included in the forced-choice task would shape participants' online interpretation of the RMET stimuli, even for adults who have years of experience with mental inference in daily social interactions. We predicted that participants in the forced-choice (i.e., traditional) version of the adult RMET would score higher on the test compared with participants who completed the free-label version. If supported, our results would replicate Cassels and Birch (2014), and provide evidence

consistent with the hypothesis that words are a context to guide participants' responses.

Agreement in the free-label version of the RMET was operationalized in two ways. First, we coded whether participants reported the target mental state word or its synonyms (defined by the Merriam-Webster online thesaurus; Merriam-Webster, 2018). Second, we used the coding scheme from Cassels and Birch (2014), assessing whether participants' responses matched the target mental state on affective valence and hostility (e.g., "positive," "negative," "neutral," or "hostile"). Cassels and Birch (2014) included "hostility" in their coding scheme based on evidence that the inability to discriminate between negative emotional states is related to other personality characteristics, such as psychopathic tendencies (for details, see Carey & Cassels, 2013). This second coding scheme allowed us to assess whether participants in the free-labeling condition offered words that were conceptually related to the target mental state word, even when their responses were not specific synonyms of the target itself.

### Method

**Participants.** We recruited 105 participants (41 male, age range 18–63,  $m = 29.48$ ,  $SD = 9.02$ ) through Amazon Mechanical Turk. One participant was disqualified from analysis for lack of compliance (i.e., failure to correctly answer the "test" question, as described below), bringing the final participant count to 104. Our sample size is approximately the same as is found in prior studies, such as the sample of typically developing participants recruited by Cassels and Birch (2014; Study 1, 118 participants). Further, a priori power analyses in G\*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) confirmed that our sample size was sufficient to detect a between-groups difference with an effect size (Cohen's  $d = .80$ ), which is smaller than that observed in Cassels and Birch (2014; Study 1,  $d = 1.67$ ) at  $\alpha < .05$  and power  $> .95$ . See Table S1 for age and gender breakdown of participants by condition. Northeastern University's Institutional Review Board approved all aspects of the study and informed consent was obtained from all participants.

**Materials and procedure.** Participants were presented with 36 cropped pictures of the eyes of an individual posing a facial configuration for a mental state (the RMET; Baron-Cohen et al., 2001), and their task was to indicate the mental state of the individual. The visual stimuli were black and white cropped photographs that showed only a pair of human eyes, eyebrows, and bridge of the nose (for sample stimuli, visit <https://www.affective-science.org/>). The following definition of mental states was provided as a reminder to all participants on each trial: *Mental states can be used to describe how a person is feeling, their attitudes toward something, or what they are thinking.* In the forced-choice condition, we followed the standard administration instructions for the RMET: on each trial, a single photograph was presented along with four mental state words and the participant was instructed to select the word that best describes the person's mental state (the target and three foils, as specified by the original test items; Baron-Cohen et al., 2001). In the free-label condition, participants were instructed to type in one word that best describes the person's mental state. We ensured that participants were actively attending to the online task, and not simply clicking through it, by randomly presenting them with a test question to which they were to select or type the response "B." All 36 trials were individually randomized for



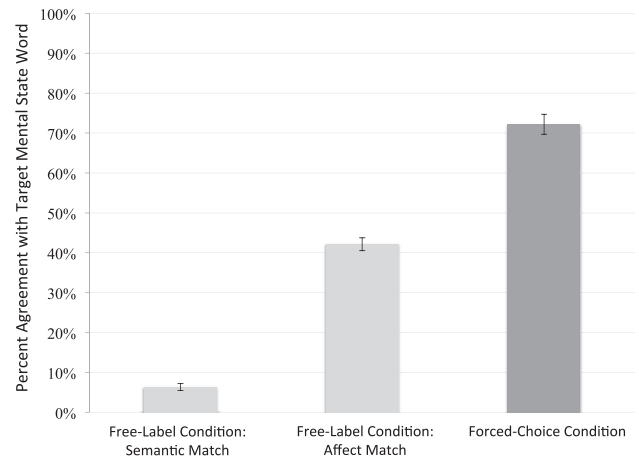
each participant. At the end of the survey, participants completed a brief demographic survey (age, ethnicity, race, gender) and were debriefed.

**Data coding.** Two researchers independently coded all responses for the free-label condition using two coding schemes. First, we used a synonym-based coding scheme to identify whether responses included either the target mental state word itself or its synonyms (defined by the Merriam-Webster online thesaurus). Responses were coded as a match (1) if they agreed with the target mental state or a mismatch (0) if they described another mental state category or behavior. See Table S2 in the Supplemental Materials for a list of terms coded as match and mismatch for all target stimuli. Interrater reliability was high: Cohen's  $\kappa = 0.84$  ( $p < .01$ ), 95% CI [0.79, 0.89]. Second, we used a valence-based coding scheme to identify whether responses matched the target mental state in terms of affective valence or hostility: positivity, negativity, neutrality, or hostility (e.g., Cassels & Birch, 2014). Responses were coded as a match (1) if they agreed with the affective category of the target mental state or a mismatch (0) if they did not agree with the affective category of the target mental state. Interrater reliability was high: Cohen's  $\kappa = .83$  ( $p < .001$ ), 95% CI [0.81, 0.85]. Discrepancies in coding were resolved by review and discussion before data were submitted for analysis.

**Statistical analysis.** We analyzed the effect of task condition on participant performance using hierarchical generalized linear modeling (HGLM; Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004). HGLM is more appropriate for binomial or categorical response data than traditional parametric approaches (e.g., independent samples  $t$  tests, ANOVA; Agresti, 2002; Jaeger, 2008), allowing us to simultaneously model both the within- and between subject variance (Guo & Zhao, 2000; Kenny, Korchmaros, & Bolger, 2003). Specifically, data were analyzed using a two-level Bernoulli model. Level 1 corresponded to trials (1 = correct response; 0 = incorrect response) that were nested within individuals (level 2) who are assigned to either the forced-choice or free-label task condition. A dummy code was used to indicate task condition. Data from the two free-label coding procedures were analyzed in separate HGLMs. The intercept represented performance in the reference free-label condition, and the  $b$  coefficient of the forced-choice term represented the difference in log-odds between conditions, with the sign indicating directionality. If, for example, the coefficient for the forced-choice term were positive, this would indicate an increase in the probability of agreement with the target mental states in this condition relative to the free-label condition. We report effect sizes and corresponding confidence intervals using the odds ratio ( $OR$ ; Fleiss & Berlin, 2009). All HGLM analyses were conducted in HLM7 (SSI Inc., Lincolnwood, IL). Model specifications are provided in online supplemental materials.

## Results

As predicted, both models comparing overall performance between task conditions indicated that agreement in the forced-choice condition was significantly greater than in the free-label condition. Participants were more likely to choose the target mental state word from a small set of alternatives than they were to freely generate the target word or its synonyms,  $b = 3.411$ ,  $SE = .101$ ,  $t(102) = 33.633$ ,  $p < .001$ ,  $OR = 30.301$ , 95% CI [24.778, 37.055]. They were also more likely to choose the target mental state word than they were to freely generate a word that matched



**Figure 1.** Mean agreement with target mental state word in the Reading the Mind in the Eyes Test (RMET) in the forced-choice versus free-label task conditions. Performance in the free-label condition is reported in two ways: based on matching either semantic meaning or affective valence of the target mental state. Error bars represent 95% CIs; these are provided as informational only.<sup>2</sup> All comparisons significant at  $p < .001$ .

the target in terms of affective valence,  $b = 1.166$ ,  $SE = .076$ ,  $t(102) = 15.337$ ,  $p < .001$ ,  $OR = 3.208$ , 95% CI [2.759, 3.730] (see Figure 1). The level of agreement observed in the forced-choice condition is consistent with previous studies using the RMET (e.g., Baron-Cohen et al., 2001).<sup>1</sup>

## Discussion

In Study 1, we found support for our hypothesis that participants would perform better in the forced-choice version of the adult RMET task that included mental state words, compared with a free-label version of the task that did not include mental state words.<sup>3</sup> These findings are consistent with the interpretation that mental state words served as a context to guide participants toward interpreting the stimuli as expected. Our findings replicated and extended those reported in Cassels and Birch (2014), as well as prior studies showing that presenting a small list of mental state

<sup>1</sup> We implemented a Bernoulli trials model for our data, in which responses are either correct (1) or incorrect (0), and chance-level responding is 50%. For this reason, we were not able to meaningfully compare participant performance against the appropriate chance levels for the current studies (i.e., 25% for the RMET stimuli in Study 1, given four response options in the forced-choice task condition; 6.25% for the emoji in Studies 2 and 3, given 16 response options).

<sup>2</sup> HGLM is not dependent on distributional variance because it compares the probability of success in binomially distributed data for two task conditions.

<sup>3</sup> Adults performed better on the RMET when compared with children (see Cassels & Birch, 2014 for data on children), both in the forced-choice version (71% of participant responses agreed with the target mental state in Study 1 versus 49% in Cassels & Birch, 2014) and in the free-labeling condition with responses coded for affective valence (41% agreement in Study 1 versus 29% in Cassels & Birch, 2014). These findings are consistent with previous published works demonstrating increased ability to make mental inferences as a function of age (Happé, Winner, & Brownell, 1998; Peterson & Slaughter, 2009) and the acquisition of specific mental state concepts (Widen & Russell, 2003).

words on each trial dramatically increases participants' agreement when making inferences based on posed facial configurations (e.g., Gendron et al., 2014b; for reviews, see Barrett et al., 2018; Gendron et al., 2018). Importantly, performance on the forced-choice version of the RMET was higher than performance on the free-label version regardless of whether we coded participants' word choices for their specific mental state content or their broader affective meaning. This finding suggests that the words embedded in the forced-choice task shaped not only the semantic interpretation of the RMET stimuli, but also their more basic, affective interpretation.

These findings have implications for forced-choice tests of mental inference, whether they are used in scientific experiments or in clinical settings, and suggest that interpretations of previous results be revisited. Critically, the purported deficits in mental inference that have been observed for many clinical disorders, including autism spectrum disorder (ASD; Baron-Cohen et al., 2001), might actually indicate a deficit in using concepts for mental states, even when cued to do so. This deficit may result from impoverished acquisition of such concepts (e.g., associated with deficits in emotion vocabulary), which in turn would diminish the impact of the words embedded in the task. Accordingly, there is evidence that people with clinical diagnoses who perform poorly on mental inference tasks also have impoverished emotion concepts and limited emotion vocabularies (i.e., alexithymia; Ozonoff, Pennington, & Rogers, 1991), and that performance on the RMET is closely related to vocabulary size (Olderbak et al., 2015). In particular, poor performance on the RMET is better predicted by alexithymia than by ASD (Oakley, Brewer, Bird, & Catmur, 2016). The current findings are in line with the hypothesis that deficits in mental inference might be reduced or even ameliorated altogether by expanding the conceptual repertoire and vocabulary for mental states.

It is also important to note that the RMET stimuli (i.e., cropped black-and-white photos of eyes) are widely used in psychological research, but may not be representative of the types of nonverbal cues that people encounter in their everyday social interactions. The images are contextually impoverished, lacking information from other facial features, bodily movements, vocal acoustics, the surrounding environment, as well as temporal information in all these channels. There is growing body of research suggesting that perceivers' social perceptions are aided by dynamic, multimodal patterns of behavior (e.g., App, McIntosh, Reed, & Hertenstein, 2011; Shuman, Clark-Polner, Meuleman, Sander, & Scherer, 2017). Presenting the images in black and white further reduces their ecological validity. In Studies 2 and 3, we built on Study 1 to examine the influence of words on mental inferences for stimuli that are more common in everyday social interactions.

## Study 2: Word Shape the Perception of Static Emoji

Emoji are pictographic symbols that portray either characters or objects (Miller et al., 2016). Since their genesis in the late 1990s, emoji use has been dramatically increasing across a range of computer-based communication platforms all over the world (Ljubešić & Fišer, 2016; for a review, see Danesi, 2016). For example, a website that monitors emoji use on Twitter—Emoji-tracker (<http://emojitracker.com>)—has detected trillions of emoji that are in common usage. The Oxford English Dictionary now

accepts emoji as a form of communication; they even chose an emoji (“laughing face with tears”) as the 2015 “Word of the Year.” Users include emoji in communication for many purposes (for reviews, see Kaye, Wall, & Malone, 2016; Kelly & Watts, 2015; Na’aman, Provenza, & Montoya, 2017), but conveying sentiment continues to be most prevalent. A worldwide sample of over a billion emoji demonstrated that the most popular emoji by far are those that communicate affective content, such as hearts, hand gestures, and faces (SwiftKey, 2015).<sup>4</sup> Face emoji (henceforth, emoji), like the emoticons that came before them, are caricatures of human facial configurations that are believed to express emotion, and are typically used to disambiguate the emotional content of a text-based message (e.g., smiling emoticons can denote a joke; Lo, 2008).<sup>5</sup>

Emoji are thought to serve as a computer-based proxy to the nonverbal gestures that humans use during face-to-face communication, but they are often subject to ambiguity and misunderstanding (Miller et al., 2016; Tigwell & Flatla, 2016). When placed in the context of an actual textual message, emoji do not always disambiguate meaning. For example, the same message with an emoji (i.e., “I miss you” with a “smiling face”) was interpreted as “sarcastic” by some participants, and “sincere” by others (Kelly & Watts, 2015). Similarly, researchers who classify emoji sentiment using information from surrounding text (without controlling for or quantifying the presence of mental state words) have found that the same emoji can be associated with opposite sentiment labels, such as the “crying face” occurring in both negative and positive contexts (Kralj Novak, Smailović, Sluban, & Mozetič, 2015). While emoji typically occur in a linguistic context, they may or may not be presented alongside emotion or other mental state words that can clarify their intended meaning. Such words may be necessary for interpreting emoji as their designers intended. For example, children with little computer experience successfully matched emoji with target emotion words when both were included in a forced-choice task (Oleszkiewicz, Frackowiak, Sorokowska, & Sorokowski, 2017).

Study 2 was designed to measure the emotional meaning of emoji that were designed to convey emotions without the need for a verbal label (Zolli, 2015).<sup>6</sup> We used emoji developed for Facebook by Pixar illustrator Matt Jones (Jones, 2017) and psychologist Dacher Keltner, who based their designs on Darwin's *The Expression of the Emotions in Man and Animals* (Darwin, 1872/2005). The emoji depict the character “Finch” as a round, yellow, freestanding face containing a mouth, eyes, eyelids, eyebrows, and wrinkles. The Finch sticker set poses 16 facial configurations (<https://www.affective-science.org/>). Because this emoji set is

<sup>4</sup> Perhaps it is for this reason that “emoji” is often thought to be etymologically related to “emotion.” The term comes instead from the Japanese characters meaning “picture + letter/character.”

<sup>5</sup> Emoji are pictographic symbols that portray characters or objects (Miller et al., 2016), whereas emoticons are “typographic symbols that appear sideways as resembling facial expressions” (Walther & d’Addario, 2001). Emoticons have been shown to successfully increase perceived emotionality of messages, and may help alleviate the feelings of psychological distance that can occur alongside computer-mediated communication (Lo, 2008). It is unclear whether emoji serve the same function.

<sup>6</sup> Although agreement levels for the prototype designs were gathered in both the United States and China, to date these data have not been published as part of peer-reviewed, scientific studies (Zolli, 2015).

based on Darwin's photographs and descriptions, it should presumably have clearer signal value than other frequently used sets, making it an ideal test case for emotion perception.<sup>7</sup> If, as predicted, we observe that participants in a forced-choice task condition (which includes a set of emotion words) have higher agreement with the target emotion than participants in a free-label task condition (which does not include those words), then this will be evidence that words shape participants' online interpretation of emoji. Given that a core goal of emoji is to help communicators avoid ambiguity (i.e., by enhancing intended predictions), quantifying the baseline (i.e., context-free) agreement for emoji is also of interest, as it can help us better understand the robustness of the high contextual variability observed in other studies to date (e.g., Kelly & Watts, 2015; Miller et al., 2016).

## Method

**Participants.** We recruited 245 U.S. participants (95 male, age range 18–65,  $m = 29.84$ ,  $SD = 9.49$ ) through Amazon Mechanical Turk. There were no significant differences in the number of men and women in each condition (see Table S3 for age and gender breakdown of participants across conditions), although condition assignment was randomized and therefore the gender and age breakdown was not strictly equal across conditions. Prior work investigating the effects of words on emotion perception has typically tested far fewer participants (e.g., Gendron et al., 2012 tested 60 participants in Study 1 and 48 participants in Study 2), suggesting that we had sufficient power to test our hypothesis. With 245 participants, we collected data for 3920 trials, comparable to the number of trials on previous studies investigating the role of context (e.g., Gendron et al., 2012: 2,880 in Study 1; 4,608 in Study 2). A priori power analyses in G\*Power 3.1 confirmed that this sample size is sufficient to detect a between-groups difference with a medium effect size (Cohen's  $d = .50$ ;  $OR = 2.50$ ) at  $\alpha < .05$  and power  $> .95$ . Two participants in the free-label condition were excluded from analysis for compliance issues (none of their responses matched the target emotion), bringing the total number of participants to 243. Northeastern University's Institutional Review Board approved all aspects of the study and informed consent was obtained from all participants.

**Materials.** We used all 16 emoji available in the Finch Sticker set as stimuli. Four of the 16 emoji from the Finch sticker set contained embedded symbolic context: tears in the sadness emoji, hearts in the love emoji, a hand scratching the head of the confusion emoji, and a red face with steam coming out of the ears for the anger emoji (see <https://www.affective-science.org/> for depictions of symbolic emoji). We omitted these four emoji from our main analyses because they are not comparable with the remaining 12 emoji, in that their additional symbolic context limits our ability to isolate the role of words as a context for emotion perception. We did, however, run additional analyses to examine the effect of words on agreement for the four symbolic emoji: see pages 13–14 and Figure S1 of the Supplemental Materials.

**Procedure.** Participants were randomly assigned to one of two task conditions in which they were asked to infer the emotional meaning of the 16 target emoji. On each trial of the forced-choice condition, participants were presented with an emoji alongside a list of 16 emotion words (a target word that matched the intended emotional meaning of the emoji and 15 foils; the words

were admiration, amusement, anger, awe, boredom, confusion, disgust, embarrassment, excitement, gratitude, happiness, love, sadness, shyness, surprise, and sympathy). Participants selected their response by clicking on the word that best matched the emoji. On each trial of the free-label condition, participants were presented with an emoji and asked to label it by typing the emotion word that best described the emotion portrayed. All 16 trials were individually randomized for each participant. At the end of the experiment, participants completed a brief demographic survey (age, ethnicity, race, gender) and were debriefed.

**Data coding.** Agreement in the free-label condition was operationalized in two ways: first, based on a semantic match, or the production of either the target emotion word or a word synonymous with the target emotion; second, based on a broader conceptual match to the target emotion.

For the semantic coding approach, two researchers first independently categorized all responses into 32 categories that included the 16 emotion categories represented by the emoji, 16 other categories for commonly generated words for which no Finch emoji existed at the time of the study (e.g., fear, disappointment, and shame), as well as categories for behavioral or situational descriptors. See Tables S6 and S7 for all 32 categories and the responses that were included in each category. Coders frequently referenced synonyms listed in the thesaurus (Merriam-Webster, 2018) when categorizing freely generated responses. After this initial categorization, participants' responses were coded as a match (1) if they agreed with the target emotion category or a mismatch (0) if they agreed with a different category. The interrater reliability was high: Cohen's  $\kappa = 0.92$  ( $p < .001$ ), 95% CI [0.90, 0.94]. Discrepancies in coding were resolved by discussion before data were submitted for analysis.

For the conceptual coding approach, one researcher coded responses as a match if the expected mental state word was listed as *either* a synonym or as a "related word" within the response's Merriam-Webster online thesaurus entry (Merriam-Webster, 2018). Although "related words: are not semantically interchangeable (i.e., synonymous) with the target emotion word, they are conceptually related. For example, "surprise" is not a synonym of "awe" (the target emotion), but it is listed as a related word. By including "related words" in this coding approach, we relaxed our threshold for a matching response to include broader categories of responses that were conceptually related to the target emotion, rather than limited to researcher- and thesaurus-defined synonyms. Participant responses that were not in the thesaurus (e.g., slang words, phrases) were removed prior to analysis, resulting in the removal of 90 trials and a reduced sample size of 240.

**Statistical analysis.** The data were analyzed using HGLM as described in Study 1. For both coding procedures, we ran models containing only the level-2 intercept and condition dummy variable to assess the overall impact of words across all emoji. Model

<sup>7</sup> The target mental states in the Finch sticker set shares considerable overlap with 10 of the 16 vocal expressions that Cordaro and colleagues (2016) have claimed to be universal. Similarly, 14 of these 16 target mental states are included in a set of 27 emotion categories that Cowen and Keltner (2017) have derived from human experience. Combined, only one of the target mental states in the Finch sticker set (gratitude) has not been recently represented in debates of universality (although see Keltner et al., in press).



specifications are provided in online supplemental materials, along with additional supporting analyses.

## Results

As predicted, both models comparing overall performance between task conditions indicated that agreement in the forced-choice condition was significantly greater than in the free-label condition. Participants were more likely to choose the target emotion word from the set of 15 alternatives than they were to freely generate that word or its synonym (as measured by our semantic coding scheme),  $b = .812$ ,  $SE = .082$ ,  $t(239) = 9.940$ ,  $p < .001$ ,  $OR = 2.252$ , 95% CI [1.917, 2.645] (see Figure 2). Similarly, participants were also more likely to choose the target emotion word than they were to freely generate a word conceptually related to the target (as measured by our conceptual coding scheme),  $b = .347$ ,  $SE = .081$ ,  $t(238) = 4.278$ ,  $p < .001$ ,  $OR = 1.415$ , 95% CI [1.206, 1.661].

A subsequent model with level-1 dummy variables for the 12 emoji without embedded symbolic context revealed an interaction between task condition and emoji. As can be seen in Figure 2, agreement levels did not differ across conditions for emoji depicting disgust, happiness, and shyness, whereas agreement was significantly lower in the free-label condition for the other emoji (with the sole exception of the emoji depicting surprise, for which agreement was significantly higher in the free-label condition; see Table S8 for detailed results).

The HGLM results illustrate differences between task conditions and individual emoji in terms of consistency of agreement (i.e., what percentage of participant responses matched the target emotion word). To further examine the specificity of agreement (i.e., to what extent the same emotion word(s) were used to label multiple emoji), we constructed confusion matrices for both task conditions. The confusion matrix for the free-label condition was

created based on the results of our semantic coding approach. As can be seen in Tables S12 and S13, consistency and specificity differed greatly by emoji and by task condition. These results demonstrate the importance of assessing for specificity to fully contextualize patterns of performance. For example, although 77% of participants provided the correct label for the surprise emoji in the forced-choice condition (and 88% in the free-label condition), this label was not applied specifically: 34% of participants also labeled the awe emoji as “surprise” in the forced-choice condition (30% in the free-label condition).

## Discussion

Overall, we found support for our hypothesis that participants were significantly more likely to perceive target emotions in face emoji when they were asked to choose an emotion word from a list of provided words, than when they were shown the emoji alone and asked to spontaneously label each one. When words were included in a forced-choice task, agreement improved for eight of the 12 included emoji (61%). This result is consistent with the hypothesis that words shape perceivers’ mental inferences for nonhuman facial symbols. This pattern held regardless of whether we coded participants’ responses based on whether they were synonymous with or more generally conceptually related to the target emotion word. Our findings demonstrate that even some highly stylized, caricatured depictions of emotion can be misidentified in the absence of context (here, emotion words), replicating and extending previous findings that emoji can be ambiguous with respect to the designer’s intentions (i.e., Kelly & Watts, 2015; Miller et al., 2016; Tigwell & Flatla, 2016).

These results also largely follow what has been observed in emotion perception studies using real, posed faces (as reviewed in Barrett et al., 2018). Interestingly, the agreement levels we observed were lower than previous forced-choice emotion perception

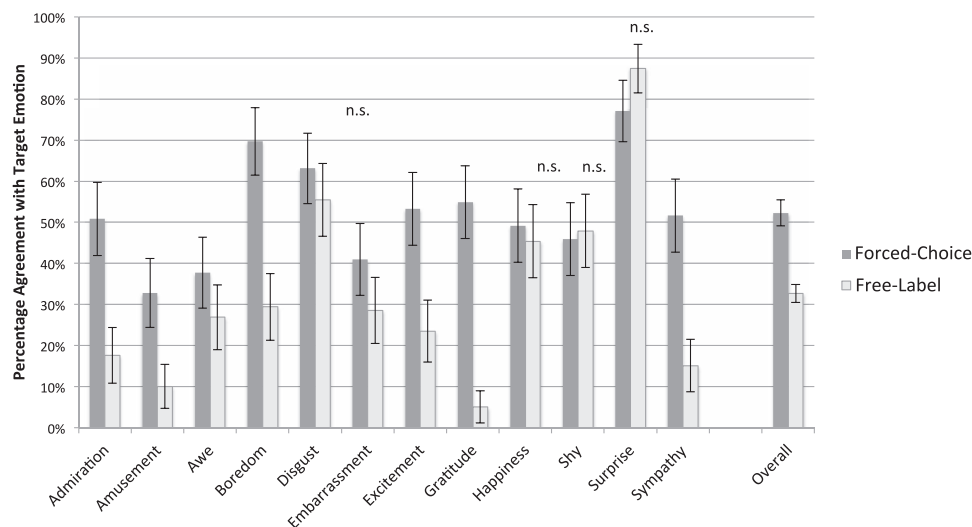


Figure 2. Mean agreement with target emotion word for 12 static “Finch” emoji in forced-choice versus free-label task conditions. Reported performance in the free-label condition is based on semantic match to target emotion. Error bars represent 95% CI; these are provided as informational only.<sup>2</sup> All between-conditions comparisons significant at  $p < .05$  unless otherwise noted: *ns* represents nonsignificance ( $p > .05$ ). Overall agreement for all 12 static emoji presented on the far right of the graph.

tasks conducted with North American English speakers (for a meta-analytic review, see Elfenbein & Ambady, 2002). One possible reason for this may be methodological: our forced-choice task included 16 response options, as compared with those that use typically use between four and six (for a review, see Russell, 1994, Table 4). More generally, the lower level of agreement suggests that these emoji may not be as clearly interpretable as originally supposed (Zolli, 2015).

It is important to note that the presence of words did not significantly improve agreement levels for the disgust, happiness, shyness, and surprise emoji. Agreement levels for the happiness and surprise emoji were equally high across conditions—in fact, agreement for the surprise emoji was actually reduced in the forced-choice condition—suggesting that certain emoji may be clearer symbols of their intended emotion categories than others. For emoji such as shyness, agreement levels did not differ between conditions but neither were they particularly high, suggesting that certain emoji may require more than the presence of the target emotion words to be fully disambiguated.

As in prior studies in which participants spontaneously labeled real, posed faces using nonmental state words (e.g., Gendron et al., 2014b; Izard, 1971; for a review, see Nelson & Russell, 2013), we also found that participants in the free-label condition sometimes did not freely generate emotion words, even when specifically instructed to do so. For example, participants labeled emoji with mental state words that are not typically considered emotions (e.g., “carefree,” “skeptical”), nonmental state words (e.g., “lively,” “sleepy”), situational descriptions (e.g., “asking permission”), and behavioral descriptions (e.g., “smirking,” “crying”). Such findings are consistent with the interpretation that free-label tasks capture the process of emotional meaning-making while imposing fewer constraints.

### Study 3: Words Shape the Perception of Dynamic Emoji

In Study 3, we replicated the methods of Study 2 using dynamic emoji in which an animated Finch portrays emotionally caricatured movements (e.g., yawning to symbolize boredom, laughing to indicate happiness). Researchers have criticized the use of static faces as ecologically valid tests of emotion perception, given that real-world facial actions are dynamic and time variant (e.g., Ambadar, Schooler, & Cohn, 2005; Caron, Caron, & Myers, 1985). In some experiments, agreement is indeed better for dynamic faces than for statically posed faces (e.g., Atkinson et al., 2004; Wallraven, Breidt, Cunningham, & Bülthoff, 2008; Wehrle, Kaiser, Schmidt, & Scherer, 2000; but also see Bould, Morris, & Wink, 2008; Kamachi et al., 2013; Widen & Russell, 2015). Therefore, in Study 3, we once again examined whether emotion words were a context for perceiving the intended emotional meaning in emoji. We predicted that participants in the forced-choice task condition would have higher agreement with the target emotion than participants in the free-label task condition, even for highly stylized emoji that contain movements as context. The stimuli used in Study 3 were dynamic versions of the static emoji used in Study 2, and therefore we also compared results from both studies to investigate whether words had similar contextual influence on emoji with and without facial movements as additional context.

## Method

**Participants.** We recruited 126 U.S. participants (47 males, age range 18–63,  $m = 29.97$ ,  $SD = 9.23$ ) through Amazon Mechanical Turk (see Table S4 for age and gender breakdown of participants by condition). A priori power analyses in G\*Power 3.1 confirmed that this sample size is sufficient to detect a between-groups difference with an effect size comparable to that observed in Study 2 ( $OR = 2.50$ ; Cohen's  $d = .50$ ) at  $\alpha < .05$  and power  $> .80$ . Four participants (two from each condition) were excluded from analysis for compliance issues (none of their responses matched the target emotion), bringing the total number of participants to 122. Northeastern University's Institutional Review Board approved all aspects of the study and informed consent was obtained from all participants.

**Materials and procedure.** We captured the 16 animated Finch emoji as they played within the Facebook messenger chat window using QuickTime Player's Screen Recording. Each animation lasted between 6 and 11 seconds and featured a moving Finch portraying a dynamic facial movement (e.g., yawning for boredom; see Table S5 for descriptions of emotionally relevant movements featured in the videos). On each of 16 trials, participants were required to press a ‘play’ button to view the emoji movie. Without pressing play, they would see the play button (an arrow) blocking the majority of the emoji. Each emoji repeated its movement twice before pausing. All other aspects of Study 3 were identical to Study 2. Like the static emoji used in Study 2, four of the dynamic emoji (anger, confusion, love, and sadness) contained embedded symbolic context. We again omitted these four emoji from our main analyses so that we could specifically test the role of words as context. See pages 18–19 and Figures S2 and S3 of Supplemental Materials for analyses of dynamic emoji with symbolic context.

**Data coding.** As in Study 2, free-label responses were coded in two ways: based on semantic relatedness and conceptual relatedness. For the semantic coding, two researchers independently coded all responses for the free-label condition following the same process reported for Study 2. The interrater reliability was high: Cohen's  $\kappa = .90$  ( $p < .01$ ), 95% CI [.88, .92]. Coding discrepancies were resolved by review and discussion before data were submitted for analysis. For the conceptual coding, one researcher coded all responses using the Merriam Webster thesaurus, following the same process reported in Study 2. Participant responses that were not in the thesaurus were removed prior to analysis, resulting in the removal of 70 trials and a reduced sample size of 121.

**Statistical analysis.** The main HGLM analyses were exactly the same for Study 3 as for Study 2. As detailed below, we also compared performance on static emoji (Study 2) against dynamic emoji (Study 3). See online supplemental materials for all model specifications and additional supporting analyses.

## Results

**Emotion perception for dynamic emoji.** Consistent with our predictions, both models comparing overall performance between task conditions indicated that agreement for the forced-choice condition was significantly greater than in the free-label condition. Replicating the results of Study 2, we found that participants were more likely to choose the target emotion word

from the set of 15 alternatives than they were to freely generate that word or its synonym (as measured by our semantic coding scheme),  $b = 1.381$ ,  $SE = .133$ ,  $t(120) = 10.408$ ,  $p < .001$ ,  $OR = 3.977$ , 95% CI [3.059, 5.172] (see Figure 3). Further, they were more likely to choose the target emotion word than they were to freely generate a word conceptually related to the target (as measured by our conceptual coding scheme),  $b = .571$ ,  $SE = .118$ ,  $t(119) = 4.839$ ,  $p < .001$ ,  $OR = 1.769$ , 95% CI [1.401, 2.235].

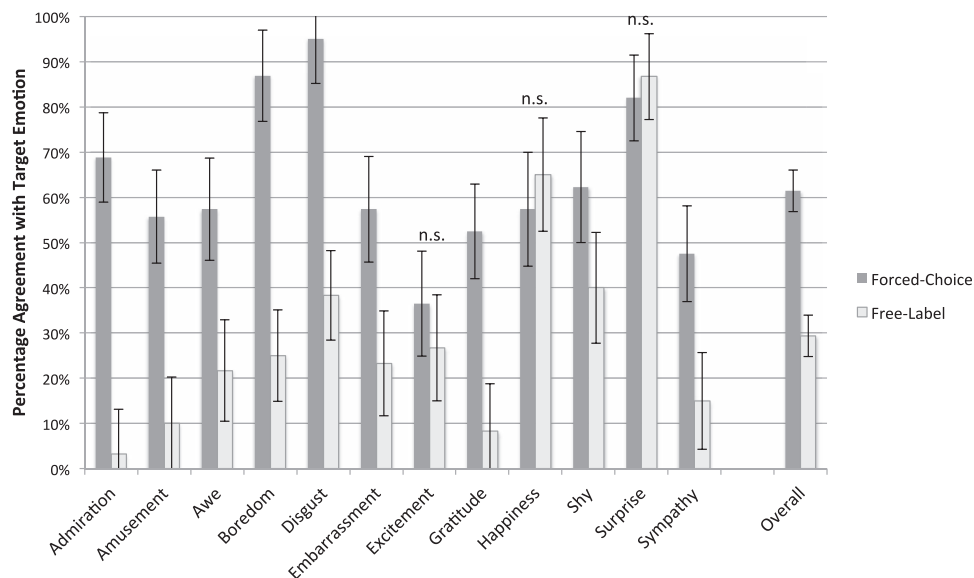
A subsequent model with level-1 dummy variables for the 12 emoji without embedded symbolic context revealed an interaction between task condition and emoji. As can be seen in Figure 3, agreement for nine of the 12 dynamic emoji was significantly lower in the free-label condition. Agreement levels did not differ across conditions for three emoji (excitement, happiness, and surprise), one of which (happiness) was also not impacted by the experimental manipulation in Study 2 (see Table S8 for detailed results). Indeed, there was some consistency in the agreement levels for individual emoji across the two studies. We observed relatively high overall agreement levels for emoji depicting boredom, disgust, and surprise. We also observed relatively low agreement levels for emoji depicting amusement, awe, and gratitude.

To compare the consistency and specificity of participant responses across task conditions and emoji, we again constructed confusion matrices. As can be seen in Tables S12 and S13, the overall pattern of performance was similar to Study 2, and indicated that labels were not necessarily applied specifically even when they were applied consistently. For example, although 100% of participants selected the correct label for the confusion emoji in the forced-choice condition, this label was also applied to other emoji by a cumulative 37% of participants.

**Emotion perception for dynamic versus static emoji.** An analysis comparing agreement levels for static and dynamic emoji allowed us to examine whether words' impact on participant agreement with target emotions differed when other context (i.e., movement) was present. We configured our level-2 model to include dummy variables for both the forced-choice condition and dynamic stimuli, as well as a Words  $\times$  Dynamic interaction term. Although the main effect of task condition remained significant across all 12 included emoji ( $b = .812$ ,  $SE = .082$ ,  $t(359) = 9.944$ ,  $p < .001$ ,  $OR = 2.252$ , 95% CI [1.918, 2.644].), the main effect of stimulus type (i.e., dynamic vs. static) was not significant ( $b = -.110$ ,  $p = .302$ ). However, these results were qualified by a significant interaction term ( $b = .569$ ,  $SE = .156$ ,  $t(359) = 3.648$ ,  $p < .001$ ,  $OR = 1.766$ , 95% CI [1.300, 2.400]), such that agreement was higher in the forced-choice condition for the dynamic emoji when compared with agreement for the static emoji (Figure S2), whereas this was not the case for the free-label condition. This finding suggests that dynamic emoji are more interpretable than static emoji when disambiguating context (i.e., an emotion word) is available, but that movement alone did not facilitate spontaneously generated labels.

## Discussion

In Study 3, we found support for our hypothesis that participants were more likely to perceive target emotions in face emoji when tested in a forced-choice task that included emotion words than when tested in a free-label task. We found that the presence of words embedded in the task increased agreement for nine of the 12 (75%) dynamic emoji. These findings are consistent with the hypothesis that words shape mental inferences even in nonhuman



**Figure 3.** Mean agreement with target emotion word for 12 dynamic “Finch” emoji in forced-choice versus free-label task conditions. Reported performance in the free-label condition is based on semantic match to target emotion. Error bars represent 95% CI; these are provided as informational only.<sup>2</sup> All between-conditions comparisons are significant at  $p < .05$  unless otherwise noted: *ns* represents nonsignificance ( $p > .05$ ). Overall agreement for all 12 dynamic emoji are presented on the far right of the graph.

facial configurations that contain other context that is designed to be disambiguating—facial movements.

Furthermore, the impact of words on emotion perception was stronger for dynamic versions of the emoji examined in Study 3 than for the static versions examined in Study 2. These findings are congruent with published studies comparing emotion perception for dynamic versus static facial stimuli which provide mixed support for a dynamic advantage (for a review, see Krumhuber, Kappas, & Manstead, 2013). Although stereotypical emotion-congruent movements have been shown to enhance emotion perception for synthetic, schematic, or degraded faces (e.g., Ehrlich, Schiano, & Sheridan, 2000; Kätsyri & Sams, 2008; Wallraven et al., 2008; Wehrle et al., 2000), this boost does not always hold for natural faces (e.g., Ehrlich et al., 2000; Fiorentini & Viviani, 2011; Kamachi et al., 2013; Kätsyri & Sams, 2008). Consistent with the present findings, studies comparing agreement levels for dynamic versus static faces using free-label tasks with children have found no dynamic advantage (e.g., Widen & Russell, 2015), supporting the interpretation that additional conceptual context—such as access to emotion knowledge—is required to achieve the expected adult levels of agreement.

As in Study 2, agreement levels for excitement, happiness, and surprise emoji did not improve by providing participants with words to select from. Free-label agreement levels were high for both the happiness and surprise emoji, and were near-ceiling level for the surprise emoji across Study 2 and 3. Agreement levels for the excitement emoji did not differ between conditions, but were also—like the shyness emoji in Study 2—not particularly high. Further research is necessary to better understand this pattern of findings. One possibility is that shyness emoji (Study 2) and excitement emoji (Study 3) were sufficiently dissimilar from their respective emotion concepts (shyness and excitement) such that the relevant emotion words could not disambiguate them. For example, even when words were provided in the forced-choice condition, participants often labeled the dynamic excitement emoji as “amusement” or “happiness,” whereas the static shyness emoji was frequently labeled as “embarrassment” or “gratitude” (see confusion matrices provided in Tables S9–10 and S12–13 for details). Another possibility is that the dynamic excitement emoji and static shyness emoji are immune to the presence of words, although this interpretation is made less plausible by the fact that the agreement levels for these emoji were not particularly high.

Finally, as in Study 2, we observed that participants in the free-label condition did not always infer emotional meaning for the emoji as indicated by the fact that they did not always spontaneously generate emotion words for all emoji. Instead, responses included mental state words not typically considered emotions, as well as nonmental state words and situational or behavioral descriptions (see Study 2 discussion and Table S7 for specific examples).

### General Discussion

A growing body of research reveals that words play a central role in constructing meaning and shaping perception more generally (for a review, see Lupyan & Clark, 2015), as well as mental inferences in particular (e.g., Chanes et al., 2018; Doyle & Lindquist, 2018; Fugate et al., 2010, 2018). In three studies, we expanded evidence for the robustness and generalizability of

these findings by showing that mental state words contribute to perceivers’ ability to infer mental meaning in eyes that are widely used in psychology experiments and in clinical assessment, as well as in caricatured symbols that are often encountered in everyday life.

### Alternative Interpretations

The present studies are not without their limitations. First, it is well known that levels of agreement in free-label tasks depend on how conservatively or liberally researchers code participant responses. In an attempt to address these effects, we used a “conservative” coding threshold based on semantic similarity (e.g., synonyms) and more “liberal” thresholds based on conceptual similarity (e.g., affective relatedness in Study 1, and conceptual relatedness in Studies 2 and 3). We observed that levels of agreement for spontaneously generated labels were indeed lower when based on our more conservative coding procedures than when we used a more liberal coding scheme, but forced-choice performance was significantly higher even when compared with the latter (free-label performance with liberal coding). Such findings indicate that our conclusions are not merely an artifact of our free-label response coding scheme.

It is also possible that participants performed better in the forced-choice task conditions because they were using compensatory strategies such as process of elimination (Widen & Russell, 2013). For example, Nelson and Russell (2016a) found that when other emotion label options and facial configurations in the trial were familiar and paired (e.g., “happy” label with smiling face), children quickly learned to pair a novel facial configuration (“puffy cheeks face”) with the only remaining, unused emotion label (“pax”). For adults, there is also evidence that preceding trials contribute to the process of elimination: participants are less likely to label a frowning face as “sad” when they labeled another frowning face as “sad” in the preceding trial (40%), compared with if the preceding trial was labeled as “anger” or “disgust” (85%; DiGirolamo & Russell, 2017, Study 4). This strategy could not account for performance in Study 1, as the response options varied across trials. It is possible that participants in Studies 2 and 3 used process of elimination to make their mental inferences, although the use of this strategy has only been observed when tasks presented participants with a small set of response options to choose from, ranging from three response options when testing children (Nelson & Russell, 2016a, 2016b; Widen & Russell, 2013) to five to eight response options when testing adults (DiGirolamo & Russell, 2017). It is unclear how efficient and effective it would be with the larger number of options (16) offered to participants in Studies 2 and 3.

### Implications for Psychological Science

Our findings are important because they invite more systematic investigations in the mechanisms by which people infer mental meaning in eyes, facial configurations, body postures, and emotional symbols. By investigating emotion perception using tasks that include words and their related conceptual content, experimenters may be creating a context for emotion perception that is not necessarily representative in everyday instances of social per-



ception. Meaningful, consistent emotion perception can indeed occur in the absence of words (e.g., as demonstrated by the emoji depicting surprise in the free-label condition of Studies 2 and 3). However, words—by cueing their associated conceptual knowledge—may encourage participants to construct different or more specific mental inferences than would be the case based on stimulus features alone. Of course, words are present as context in many social interactions, but usually not organized as a small set of options for the perceiver to select from. Our findings and interpretation therefore strongly suggest the need for future work investigating the boundary conditions on when and to what degree words shape mental inferences, not only within psychology laboratories, but during clinical assessments and social perceptions as they occur in the real world, particularly when those perceptions influence target's outcomes, such as in a courtroom or hospital.

Our findings also highlight the impact of experimental design features on estimates of mental inference. The reliance on forced-choice task designs might strengthen replication in the lab, but will weaken how generalizable the findings are to the outside world where context is less constrained and sensory inputs are more variable. The assumption that findings generalize from highly constrained experimental contexts to everyday life can have potentially hazardous consequences. For example, children on the autism spectrum are currently taught to recognize facial configurations as emotional expressions (Baron-Cohen, Golan, Wheelwright, & Hill, 2004; Kouo & Egel, 2016), even though these configurations are not produced in a consistent and specific manner (for a review, see Barrett et al., 2018), and the consistency and specificity of neurotypical perceivers' responses is increased by the presence of words or other conceptual content (e.g., Study 1; Cassels & Birch, 2014). The present findings thus substantiate concerns with using forced-choice tasks to assess emotion perception and mental inference (e.g., Barrett et al., 2018; Hoemann et al., in press; Russell, 1994), suggesting the need for converging evidence from multiple and ecologically valid methods when assessing reliability and application of scientific findings.

In underscoring the role of words and other conceptual content in shaping mental inference, the present findings are consistent with emerging predictive coding accounts of mental life, including social perception. Emerging evidence in neuroscience indicates that the human brain uses conceptual knowledge as a top-down, Bayesian filter to anticipate and make sense of sensory inputs to guide actions (see Barrett, 2017a, 2017b; Barrett & Simmons, 2015; Clark, 2013; Hohwy, 2013; Friston, 2010). The brain is thought to construct prediction signals at multiple time scales and levels of specificity, including specialized perceptual levels related to real-time sampling of the world (Kiebel, Daunizeau, & Friston, 2008). This predictive coding account has been proposed as a general theory of perception and cognition (Barrett, 2017a, 2017b; K. Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2017; Spratling, 2016, 2017), action (Wolpert & Flanagan, 2001), language (Lupyan & Clark, 2015), mood and affect (Barrett, Quigley, & Hamilton, 2016; Barrett & Simmons, 2015), and consciousness (Chanes & Barrett, 2016). Predictive coding has further been extended to social perception (Otten, Seth, & Pinto, 2017), including mental inference (Koster-Hale & Saxe, 2013; Thornton & Tamir, 2017) and emotion perception (Barrett, 2017b; Chanes et al., 2018). Recent evidence indicates that expectations constrain initial perceptions rather than being downstream products of social

perception (Freeman & Johnson, 2016). Correspondingly, the words in forced-choice tasks, particularly when they are repeated from trial to trial, might implicitly shape the representations that serve as Bayesian filters on subsequent trials, constraining and shaping mental inferences. This speculative hypothesis awaits future testing.

## Implications for Communication

The current findings may also have practical implications for daily communication. The lack of consistency in emoji interpretation has led researchers to develop classifiers of emoji meaning using the sentiment of the surrounding text (e.g., Liu, Li, & Guo, 2012; Wijeratne, Balasuriya, Sheth, & Doran, 2016), and has led users to create reference materials such as an online emojiopedia (Emojiopedia, 2018). Indeed, emoji-only messages are rare, because of the high variability and difficulty of interpretation (Danesi, 2016). Part of this difficulty may be because of the very nature of emoji: although they are designed to replicate (involuntary) non-verbal facial expressions, they—like words—are employed as “deliberately encoded elements of intentional communication” (Derks, Bos, & von Grumbkow, 2007, p. 847; Walther & D’Addario, 2001). Indeed, the same emoji can be used to convey variable meanings that the receiver must decipher based on context (e.g., a winking face could denote a joke or flirting). Future studies could expand upon the current findings by investigating how context surrounding emoji is created during naturalistic interactions, and the potential causes and downstream consequences of emoji misinterpretation.

The global surge in emoji popularity has also led to different uses and interpretations by culture (for a review, see Danesi, 2016), and has even inspired businesses to recruit culturally sensitive translations specialists for “the world’s fastest growing language” (“Emoji Translator/Specialist,” 2017). When comparing cross-cultural emotion perception for “happy,” “neutral,” and “sad” photos of faces, emoticons, and black-and-white emoji, researchers recently found that emotion was not universally perceived in the latter two stimulus types, but rather varied by exposure (Takahashi, Oishi, & Shimada, 2017). These results corroborate the present findings that context can be critical for purportedly international symbols, and that context extends beyond the stimulus and its immediate linguistic or experimental environment to include individuals’ communicative habits and cultural values. Future studies on computer-mediated communication could formally compare how users from different cultures select and use emoji, as well as other culturally relative Internet expressions such as GIFs (Hess & Bui, 2017).

## References

- Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). New York, NY: Wiley. <http://dx.doi.org/10.1002/0471249688>
- Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, 16, 403–410. <http://dx.doi.org/10.1111/j.0956-7976.2005.01548.x>
- App, B., McIntosh, D. N., Reed, C. L., & Hertenstein, M. J. (2011). Nonverbal channel use in communication of emotion: How may depend on why. *Emotion*, 11, 603–617. <http://dx.doi.org/10.1037/a0023164>

- Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, 33, 717–746. <http://dx.doi.org/10.1068/p5096>
- Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., . . . Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, 19, 724–732. <http://dx.doi.org/10.1111/j.1467-9280.2008.02148.x>
- Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, 338, 1225–1229. <http://dx.doi.org/10.1126/science.1224313>
- Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology*, 13, 379–398. <http://dx.doi.org/10.1111/j.2044-835X.1995.tb00687.x>
- Baron-Cohen, S., & Cross, P. (1992). Reading the eyes: Evidence for the role of perception in the development of a theory of mind. *Mind & Language*, 7, 172–186. <http://dx.doi.org/10.1111/j.1468-0017.1992.tb00203.x>
- Baron-Cohen, S., Golan, O., Wheelwright, S., & Hill, J. J. (2004). *Mind reading: The interactive guide to emotions*. London, UK: Jessica Kingsley Limited.
- Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology & Psychiatry, & Allied Disciplines*, 38, 813–822. <http://dx.doi.org/10.1111/j.1469-7610.1997.tb01599.x>
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 241–251. <http://dx.doi.org/10.1111/1469-7610.00715>
- Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a “language of the eyes”? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition*, 4, 311–331. <http://dx.doi.org/10.1080/713756761>
- Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, 10, 20–46. [http://dx.doi.org/10.1207/s15327957pspr1001\\_2](http://dx.doi.org/10.1207/s15327957pspr1001_2)
- Barrett, L. F. (2011). Was Darwin wrong about emotional expressions? *Current Directions in Psychological Science*, 20, 400–406. <http://dx.doi.org/10.1177/0963721411429125>
- Barrett, L. F. (2017a). *How emotions are made: The secret life the brain and what it means for your health, the law, and human nature*. New York, NY: Houghton Mifflin Harcourt.
- Barrett, L. F. (2017b). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12, 1833–1855. <http://dx.doi.org/10.1093/scan/nsx060>
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A., & Pollak, S. (2018). *Emotional expressions reconsidered: Challenges to inferring emotion in human facial movements*. Manuscript submitted for publication.
- Barrett, L. F., & Gendron, M. (2016). The importance of context: Three corrections to Cordaro, Keltner, Tshering, Wangchuk, and Flynn (2016). *Emotion*, 16, 803–806. <http://dx.doi.org/10.1037/emo0000217>
- Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, 11, 327–332. <http://dx.doi.org/10.1016/j.tics.2007.06.003>
- Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, 20, 286–290. <http://dx.doi.org/10.1177/0963721411422522>
- Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, 371, 20160011.
- Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, 16, 419–429. <http://dx.doi.org/10.1038/nrn3950>
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19, 2767–2796. <http://dx.doi.org/10.1093/cercor/bhp055>
- Boucher, J. D., & Carlson, G. E. (1980). Recognition of facial expression in three cultures. *Journal of Cross-Cultural Psychology*, 11, 263–280. <http://dx.doi.org/10.1177/0022022180113003>
- Bould, E., Morris, N., & Wink, B. (2008). Recognising subtle emotional expressions: The role of facial movements. *Cognition and Emotion*, 22, 1569–1587. <http://dx.doi.org/10.1080/02699930801921156>
- Brooks, J. A., Shablack, H., Gendron, M., Satpute, A. B., Parrish, M. H., & Lindquist, K. A. (2017). The role of language in the experience and perception of emotion: A neuroimaging meta-analysis. *Social Cognitive and Affective Neuroscience*, 12, 169–183.
- Carey, J. M., & Cassels, T. G. (2013). Comparing two forms of a childhood perspective-taking measure using CFA and IRT. *Psychological Assessment*, 25, 879–892. <http://dx.doi.org/10.1037/a0032641>
- Caron, R. F., Caron, A. J., & Myers, R. S. (1985). Do infants see emotional expressions in static faces? *Child Development*, 56, 1552–1560. <http://dx.doi.org/10.2307/1130474>
- Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*, 70, 205–218. <http://dx.doi.org/10.1037/0022-3514.70.2.205>
- Cassels, T. G., & Birch, S. A. (2014). Comparisons of an open-ended vs. forced-choice ‘mind reading’ task: Implications for measuring perspective-taking and emotion recognition. *PLoS ONE*, 9(12), e93653. <http://dx.doi.org/10.1371/journal.pone.0093653>
- Chanes, L., & Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences*, 20, 96–106. <http://dx.doi.org/10.1016/j.tics.2015.11.005>
- Chanes, L., Wormwood, J. B., Betz, N., & Barrett, L. F. (2018). Facial expression predictions as drivers of social perception. *Journal of Personality and Social Psychology*, 114, 380–396. <http://dx.doi.org/10.1037/pspa0000108>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181–204. <http://dx.doi.org/10.1017/S0140525X12000477>
- Clark-Polner, E., Wager, T. D., Satpute, A. B., & Barrett, L. F. (2016). Neural fingerprinting: Meta-analysis, variation, and the search for brain-based essences in the science of emotion. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *The handbook of emotion* (4th ed., pp. 146–165). New York, NY: Guilford Press.
- Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion*, 16, 117–128. <http://dx.doi.org/10.1037/emo0000100>
- Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E7900–E7909. <http://dx.doi.org/10.1073/pnas.1702247114>
- Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J. M. (2016). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General*, 145, 830–843. <http://dx.doi.org/10.1037/xge0000172>
- Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J.-M. (2016). The fear gasping face as a threat display in a Melanesian society. *Proceedings of the National Academy of Sciences of the United States*

- of America, 113, 12403–12407. <http://dx.doi.org/10.1073/pnas.1611622113>
- Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J.-M. (2017). Recognizing spontaneous facial expressions of emotion in a small-scale society of Papua New Guinea. *Emotion*, 17, 337–347. <http://dx.doi.org/10.1037/emo0000236>
- Danesi, M. (2016). *The semiotics of emoji: The rise of visual language in the age of the internet*. London, UK: Bloomsbury Publishing.
- Darwin, C. (2005). *The expression of the emotions in man and animals*: Digireads.com Publishing. <http://dx.doi.org/10.1037/10001-000> (Original work published 1872)
- de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, 7, 242–249. <http://dx.doi.org/10.1038/nrn1872>
- Derks, D., Bos, A. E., & von Grumbkow, J. (2007). Emoticons and social interaction on the Internet: The importance of social context. *Computers in Human Behavior*, 23, 842–849. <http://dx.doi.org/10.1016/j.chb.2004.11.013>
- DiGirolamo, M. A., & Russell, J. A. (2017). The emotion seen in a face can be a methodological artifact: The process of elimination hypothesis. *Emotion*, 17, 538–546. <http://dx.doi.org/10.1037/emo0000247>
- Doyle, C. M., & Lindquist, K. A. (2018). When a word is worth a thousand pictures: Language shapes perceptual memory for emotion. *Journal of Experimental Psychology: General*, 147, 62–73. <http://dx.doi.org/10.1037/xge0000361>
- Dyck, M. J., & Denver, E. (2003). Can the emotion recognition ability of deaf children be enhanced? A pilot study. *Journal of Deaf Studies and Deaf Education*, 8, 348–356. <http://dx.doi.org/10.1093/deafed/eng019>
- 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, e3628. <http://dx.doi.org/10.1371/journal.pone.0003628>
- Ehrlich, S. M., Schiano, D. J., & Sheridan, K. (2000). Communicating facial affect: It's not the realism, it's the motion. *CHI'00 extended abstracts on human factors in computing systems* (pp. 251–252). The Hague, the Netherlands: ACM.
- Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion*, 10, 159–168. <http://dx.doi.org/10.1007/BF00992253>
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128, 203–235. <http://dx.doi.org/10.1037/0033-2909.128.2.203>
- Emojipedia. (2018). Retrieved March 5, 2018 <https://emojipedia.org/>
- Emoji Trans./Specialist. (2017). Retrieved from <https://www.todaytranslations.com/emoji-Trans.-specialist>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160. <http://dx.doi.org/10.3758/BRM.41.4.1149>
- Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *Journal of Vision*, 11, 17–17. <http://dx.doi.org/10.1167/11.3.17>
- Fleiss, J. L., & Berlin, J. A. (2009). Effect sizes for dichotomous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 237–253). New York, NY: Russell Sage Foundation.
- Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences*, 20, 362–374. <http://dx.doi.org/10.1016/j.tics.2016.03.003>
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11, 127–138. <http://dx.doi.org/10.1038/nrn2787>
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, 29, 1–49. [http://dx.doi.org/10.1162/NECO\\_a\\_00912](http://dx.doi.org/10.1162/NECO_a_00912)
- Fugate, J., Gendron, M., Nakashima, S., & Barrett, L. F. (2018). Emotion words: Adding face value. *Emotion*, 18, 693–706.
- Fugate, J. M., Gouzoules, H., & Barrett, L. F. (2010). Reading chimpanzee faces: Evidence for the role of verbal labels in categorical perception of emotion. *Emotion*, 10, 544–554. <http://dx.doi.org/10.1037/a0019017>
- Gelman, S. A., & Roberts, S. O. (2017). How language shapes the cultural inheritance of categories. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 7900–7907. <http://dx.doi.org/10.1073/pnas.1621073114>
- Gendron, M., Crivelli, C., & Barrett, L. F. (2018). Universality Reconsidered: Diversity in Making Meaning of Facial Expressions. *Current Directions in Psychological Science*, 27, 211–219. <http://dx.doi.org/10.1177/0963721417746794>
- Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, 12, 314–325. <http://dx.doi.org/10.1037/a0026007>
- Gendron, M., Mesquita, B., & Barrett, L. F. (2013). Emotion perception: Putting the face in context. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 379–389). New York, NY: Oxford University Press.
- Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, 25, 911–920. <http://dx.doi.org/10.1177/0956797613517239>
- Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014b). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*, 14, 251–262. <http://dx.doi.org/10.1037/a0036052>
- Guo, G., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology*, 26, 441–462. <http://dx.doi.org/10.1146/annurev.soc.26.1.441>
- Halberstadt, J. B., & Niedenthal, P. M. (2001). Effects of emotion concepts on perceptual memory for emotional expressions. *Journal of Personality and Social Psychology*, 81, 587–598. <http://dx.doi.org/10.1037/0022-3514.81.4.587>
- Happé, F. G., Winner, E., & Brownell, H. (1998). The getting of wisdom: Theory of mind in old age. *Developmental Psychology*, 34, 358–362. <http://dx.doi.org/10.1037/0012-1649.34.2.358>
- Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently ambiguous: Facial expressions of emotions, in context. *Emotion Review*, 5, 60–65. <http://dx.doi.org/10.1177/1754073912451331>
- Hess, A., & Bui, Q. (2017). *What love and sadness look like in 5 countries, according to their top GIFs*. Retrieved from <https://www.nytimes.com/interactive/2017/12/29/upshot/gifs-emotions-by-country.html?smid=fb-share>
- Hoemann, K., Crittenden, A. N., Msafiri, S., Liu, Q., Li, C., Roberson, D., . . . Barrett, L. F. (in press). Context facilitates performance on a classic cross-cultural emotion perception task. *Emotion*.
- Hohwy, J. (2013). *The predictive mind*. Unite Kingdom: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199682737.001.0001>
- Izard, C. E. (1971). *The face of emotion*. New York, NY: Appleton-Century-Crofts.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446. <http://dx.doi.org/10.1016/j.jml.2007.11.007>
- Jones, M. (2017). *St. David's day doodle!* Retrieved March 2, 2017, from <http://mattjonezanimation.blogspot.com/>
- Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2013). Dynamic properties influence the perception of facial expressions. *Perception*, 42, 1266–1278. <http://dx.doi.org/10.1068/p3131n>
- Käsytari, J., & Sams, M. (2008). The effect of dynamics on identifying basic emotions from synthetic and natural faces. *International Journal of*



- Human-Computer Studies*, 66, 233–242. <http://dx.doi.org/10.1016/j.jhcs.2007.10.001>
- Kaye, L. K., Wall, H. J., & Malone, S. A. (2016). “Turn that frown upside-down”: A contextual account of emoticon usage on different virtual platforms. *Computers in Human Behavior*, 60, 463–467. <http://dx.doi.org/10.1016/j.chb.2016.02.088>
- Kelly, R., & Watts, L. (2015, September). *Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships*. Paper presented at the experiences of technology appropriation: Unanticipated Users, Usage, Circumstances, and Design, Oslo, Norway.
- Keltner, D. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition and Emotion*, 10, 155–172. <http://dx.doi.org/10.1080/026999396380312>
- Keltner, D., Sauter, D., Tracy, J. L., & Cowen, A. S. (in press). Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior*.
- Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, 8, 115–128. <http://dx.doi.org/10.1037/1082-989X.8.2.115>
- Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, 4, e1000209. <http://dx.doi.org/10.1371/journal.pcbi.1000209>
- Kleinsmith, A., & Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, 4, 15–33. <http://dx.doi.org/10.1109/T-AFFC.2012.16>
- Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron*, 79, 836–848. <http://dx.doi.org/10.1016/j.neuron.2013.08.020>
- Kouo, J. L., & Egel, A. L. (2016). The effectiveness of interventions in teaching emotion recognition to children with autism spectrum disorder. *Review Journal of Autism and Developmental Disorders*, 3, 254–265. <http://dx.doi.org/10.1007/s40489-016-0081-1>
- Kralj Novak, P., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of Emojis. *PLoS ONE*, 10, e0144296. <http://dx.doi.org/10.1371/journal.pone.0144296>
- Krumhuber, E. G., Kappas, A., & Manstead, A. S. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, 5, 41–46. <http://dx.doi.org/10.1177/1754073912451349>
- Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion*, 6, 125–138. <http://dx.doi.org/10.1037/1528-3542.6.1.125>
- Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion*, 14, 375–387. <http://dx.doi.org/10.1037/a0035293>
- Liu, K. L., Li, W. J., & Guo, M. (2012). Emoticon smoothed language models for twitter sentiment analysis. *Proceedings of the 26th AAAI Conference on Artificial Intelligence* (pp. 1678–1684). Palo Alto, CA: AAAI Press.
- Ljubešić, N., & Fišer, D. (2016). *A global analysis of emoji usage*. Paper presented at the Proceedings of the 10th Web as Corpus Workshop.
- Lo, S. K. (2008). The nonverbal communication functions of emoticons in computer-mediated communication. *CyberPsychology & Behavior*, 11, 595–597. <http://dx.doi.org/10.1089/cpb.2007.0132>
- Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, 24, 279–284. <http://dx.doi.org/10.1177/0963721415570732>
- Merriam-Webster. (2018). Retrieved March 5, 2018 <https://www.merriam-webster.com/>
- Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016). “Blissfully happy” or “ready to fight”: Varying interpretations of emoji. Paper presented at the 10th International Conference on Web and Social Media, ICWSM 2016.
- Na’aman, N., Provenza, H., & Montoya, O. (2017). Varying linguistic purposes of emoji in (Twitter) context. *Proceedings of ACL 2017, Student Research Workshop* (pp. 136–141).
- Nelson, N. L., & Russell, J. A. (2013). Universality revisited. *Emotion Review*, 5, 8–15. <http://dx.doi.org/10.1177/1754073912457227>
- Nelson, N. L., & Russell, J. A. (2016a). Building emotion categories: Children use a process of elimination when they encounter novel expressions. *Journal of Experimental Child Psychology*, 151, 120–130. <http://dx.doi.org/10.1016/j.jecp.2016.02.012>
- Nelson, N. L., & Russell, J. A. (2016b). A facial expression of pax: Assessing children’s “recognition” of emotion from faces. *Journal of Experimental Child Psychology*, 141, 49–64. <http://dx.doi.org/10.1016/j.jecp.2015.07.016>
- Nook, E. C., Lindquist, K. A., & Zaki, J. (2015). A new look at emotion perception: Concepts speed and shape facial emotion recognition. *Emotion*, 15, 569–578. <http://dx.doi.org/10.1037/a0039166>
- Oakley, B. F. M., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. *Journal of Abnormal Psychology*, 125, 818–823. <http://dx.doi.org/10.1037/abn0000182>
- Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: Toward a brief form for research and applied settings. *Frontiers in Psychology*, 6, 1503. <http://dx.doi.org/10.3389/fpsyg.2015.01503>
- Oleszkiewicz, A., Frackowiak, T., Sorokowska, A., & Sorokowski, P. (2017). Children can accurately recognize facial emotions from emoticons. *Computers in Human Behavior*, 76, 372–377. <http://dx.doi.org/10.1016/j.chb.2017.07.040>
- Otten, M., Seth, A. K., & Pinto, Y. (2017). A social Bayesian brain: How social knowledge can shape visual perception. *Brain and Cognition*, 112, 69–77. <http://dx.doi.org/10.1016/j.bandc.2016.05.002>
- Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology & Psychiatry, & Allied Disciplines*, 32, 1081–1105. <http://dx.doi.org/10.1111/j.1469-7610.1991.tb00351.x>
- Peterson, C. C., & Slaughter, V. (2009). Theory of mind (ToM) in children with autism or typical development: Links between eye-reading and false belief understanding. *Research in Autism Spectrum Disorders*, 3, 462–473. <http://dx.doi.org/10.1016/j.rasd.2008.09.007>
- Raudenbush, S., Bryk, A., Cheong, Y., Congdon, R., & du Toit, M. (2004). *HLM 6 hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.
- Raz, G., Touroutoglou, T., Wilson-Mendenhall, C., Gilam, G. L. T., Gonen, T., Jacob, Y., . . . Barrett, L. F. (2016). Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. *Cognitive, Affective, & Behavioral Neuroscience*, 16, 709–723.
- Richell, R., Mitchell, D., Newman, C., Leonard, A., Baron-Cohen, S., & Blair, R. (2003). Theory of mind and psychopathy: Can psychopathic individuals read the ‘language of the eyes’? *Neuropsychologia*, 41, 523–526.
- Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition*, 71, 1–42. [http://dx.doi.org/10.1016/S0010-0277\(99\)00013-X](http://dx.doi.org/10.1016/S0010-0277(99)00013-X)
- Rodrigues, S. M., Saslow, L. R., Garcia, N., John, O. P., & Keltner, D. (2009). Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 21437–21441. <http://dx.doi.org/10.1073/pnas.0909579106>



- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102–141. <http://dx.doi.org/10.1037/0033-2909.115.1.102>
- Russell, J. A. (1997). 13-reading emotion from and into faces: Resurrecting a dimensional-contextual perspective. *The psychology of facial expression*, 295–320.
- Schick, B., de Villiers, P., de Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. *Child Development*, 78, 376–396. <http://dx.doi.org/10.1111/j.1467-8624.2007.01004.x>
- Shariff, A. F., & Tracy, J. L. (2011). What are emotion expressions for? *Current Directions in Psychological Science*, 20, 395–399. <http://dx.doi.org/10.1177/0963721411424739>
- Shuman, V., Clark-Polner, E., Meuleman, B., Sander, D., & Scherer, K. R. (2017). Emotion perception from a componential perspective. *Cognition and Emotion*, 31, 47–56. <http://dx.doi.org/10.1080/02699931.2015.1075964>
- Sidera, F., Amadó, A., & Martínez, L. (2017). Influences on facial emotion recognition in deaf children. *Journal of Deaf Studies and Deaf Education*, 22, 164–177. <http://dx.doi.org/10.1093/deafed/enw072>
- Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., . . . Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin*, 144, 343–393. <http://dx.doi.org/10.1037/bul0000128>
- Spencer, P. E., & Marschark, M. (2010). *Evidence-based practice in educating deaf and hard-of-hearing students*. New York, NY: Oxford University Press.
- Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cognitive Processing*, 17, 279–305. <http://dx.doi.org/10.1007/s10339-016-0765-6>
- Spratling, M. W. (2017). A hierarchical predictive coding model of object recognition in natural images. *Cognitive Computation*, 9, 151–167. <http://dx.doi.org/10.1007/s12559-016-9445-1>
- SwiftKey. (2015). Most-used emoji revealed: Americans love skulls, Brazilians love cats, the French love hearts [Blog post]. Retrieved from <https://blog.swiftkey.com/americans-love-skulls-brazilians-love-cats-swiftkey-emoji-meanings-report/>
- Takahashi, K., Oishi, T., & Shimada, M. (2017). Is ☺ smiling? Cross-cultural study on recognition of emoticon's emotion. *Journal of Cross-Cultural Psychology*, 48, 1578–1586. <http://dx.doi.org/10.1177/0022022117734372>
- Thornton, M. A., & Tamir, D. I. (2017). Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 5982–5987. <http://dx.doi.org/10.1073/pnas.1616056114>
- Tigwell, G. W., & Flatla, D. R. (2016). Oh that's what you meant!: reducing emoji misunderstanding. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 859–866). Florence, Italy: ACM.
- 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, 741–749. <http://dx.doi.org/10.1016/j.chb.2010.10.018>
- Tonks, J., Williams, W. H., Frampton, I., Yates, P., & Slater, A. (2007). Reading emotions after child brain injury: A comparison between children with brain injury and non-injured controls. *Brain Injury*, 21, 731–739. <http://dx.doi.org/10.1080/02699050701426899>
- Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological Science*, 15, 194–197. <http://dx.doi.org/10.1111/j.0956-7976.2004.01503008.x>
- Tracy, J. L., Robins, R. W., & Schriber, R. A. (2009). Development of a FACS-verified set of basic and self-conscious emotion expressions. *Emotion*, 9, 554–559. <http://dx.doi.org/10.1037/a0015766>
- Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, 94, 3–15. <http://dx.doi.org/10.1037/0033-295X.94.1.3>
- Wallraven, C., Breidt, M., Cunningham, D. W., & Bühlhoff, H. H. (2008). Evaluating the perceptual realism of animated facial expressions. [TAP]. *ACM Transactions on Applied Perception*, 4, 1–20. <http://dx.doi.org/10.1145/1278760.1278764>
- Walther, J. B., & D'Addario, K. P. (2001). The impacts of emoticons on message interpretation in computer-mediated communication. *Social Science Computer Review*, 19, 324–347. <http://dx.doi.org/10.1177/089443930101900307>
- Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, 78, 105–119. <http://dx.doi.org/10.1037/0022-3514.78.1.105>
- Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review*, 5, 72–77. <http://dx.doi.org/10.1177/1754073912451492>
- Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, 39, 114–128. <http://dx.doi.org/10.1037/0012-1649.39.1.114>
- Widen, S. C., & Russell, J. A. (2013). Children's recognition of disgust in others. *Psychological Bulletin*, 139, 271–299. <http://dx.doi.org/10.1037/a0031640>
- Widen, S. C., & Russell, J. A. (2015). Do dynamic facial expressions convey emotions to children better than do static ones? *Journal of Cognition and Development*, 16, 802–811. <http://dx.doi.org/10.1080/15248372.2014.916295>
- 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. <http://dx.doi.org/10.3389/fpsyg.2012.00471>
- Wijeratne, S., Balasuriya, L., Sheth, A., & Doran, D. (2016). Emojinet: Building a machine readable sense inventory for emoji. *International conference on social informatics* (pp. 527–541). Cham, Switzerland/Rome, Italy: Springer.
- Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2015). Variety in emotional life: Within-category typicality of emotional experiences is associated with neural activity in large-scale brain networks. *Social Cognitive and Affective Neuroscience*, 10, 62–71. <http://dx.doi.org/10.1093/scan/nsu037>
- Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, 11, R729–R732. [http://dx.doi.org/10.1016/S0960-9822\(01\)00432-8](http://dx.doi.org/10.1016/S0960-9822(01)00432-8)
- Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. (2001). *Woodcock-Johnson R III NU Tests of Achievement*. Itasca, IL: Riverside.
- Zolli, A. (2015). *Darwin's stickers*. Retrieved from <http://www.radiolab.org/story/darwins-stickers/>

Received June 8, 2017

Revision received July 23, 2018

Accepted July 23, 2018 ■