COMMENTARY

Context may reveal how you feel

Aleix M. Martinez[a,b,1]

Face perception is a fundamental component of our cognitive system and, arguably, a core ability that allowed humans to create the large, advanced societies of today. When we look at someone else's face, we recognize who they are, whether they are female or male, attractive or unattractive, and happy or sad; that is, their affective state. Correctly interpreting these signals is essential for a functional, cooperative society. For example, when looking at the faces in Fig. 1, most people identify a female expressing sadness on the left and an angry male on the right. But while identity and other attributes are recognized quite accurately (1), affect is not (2). To see this, look at the images in Fig. 2*A* and *B*. What expressions would you now say these two individuals express? Most of us classify them as expressing excitement or euphoria; that is, positive emotions. What is behind this radical change in our interpretation of these images? Context. Our interpretation of a facial configuration is dependent on the context in which the facial expression is situated. In an ambitious new study in PNAS, Chen and Whitney (3) show that people make reasonably good predictions of people's affect when only the contextual information is known; that is, when the face is not observable (Fig. 2*C*). This inference is shown to be accurate, even when the whole body of the person is masked (Fig. 2*D*), thus preventing an inference based on body pose. Context, therefore, is not only necessary for a correct interpretation of how others feel but, in some instances, it is sufficient. This surprising result will provide renewed interest in the value that context plays in our interpretation of how others feel.

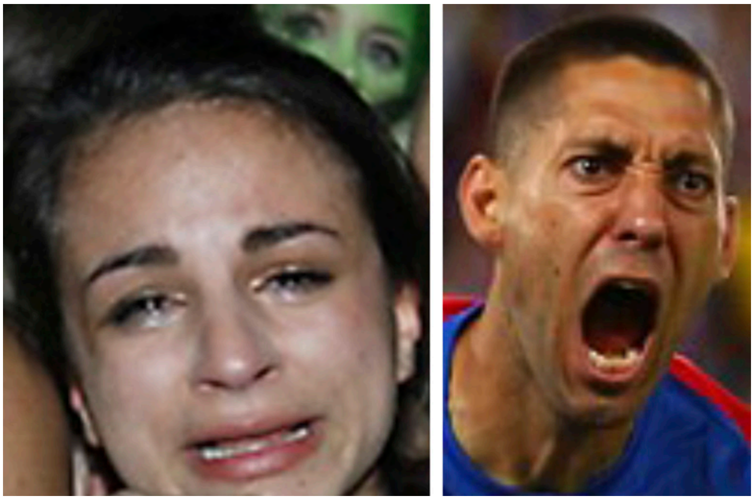

Fig. 1. **When asked to identify the emotions shown in these images, most people agree that the left image expresses sadness, while the right image is a clear display of anger. If asked whether these expressions communicate positive or negative valence, most people agree that both correspond to a negative expression. The problem with these assessments is that context is not observable, which may lead to incorrect interpretations. Images courtesy of (*Left*) Imgflip and (*Right*) Getty Images/Michael Steele.**

But why do people generally believe that the face alone is the necessary and sufficient visual cue to interpret affect? To better understand this, consider the following: When we find ourselves in a specific situation, our central nervous system executes a number of computations intended to keep us from harm and to maximize the likelihood of achieving our goals. Some of these computations yield facial muscle movements that are observable to friends and foes. Because of this, we tend to assume that we can always interpret what a person is feeling (i.e., their affect) by visually examining their facial configuration (4). This interpretation of emotive expressions assumes that the facial articulations produced in each situation are unique. However, as the examples in Figs. 1 and 2 demonstrate, this is not always the case. A facial configuration can be correctly interpreted only when its context is observed or known. While it is true that the female in Fig. 1 seems to express sadness and that her facial configuration is typically seen when someone is sad, the same facial configuration may be produced in a variety of other situations, like when a fan finds herself face-to-face with her idol (Fig. 2*A*).

[a]Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210; and [b]Center for Cognitive and Brain Sciences, The Ohio State University, Columbus, OH 43210

Author contributions: A.M.M. wrote the paper.

Conflict of interest statement: The author and The Ohio State University hold several patents on algorithms for the automatic recognition of facial expressions of emotion.

See companion article 10.1073/pnas.1812250116.

[1]Email: martinez.158@osu.edu.

Fig. 2. Adding context to the facial expressions previously seen in Fig. 1 radically changes our interpretation of the emotion being experienced by a person. (*A* and *B*) In these two images, most observers agree that the people shown are experiencing a joyful event (i.e., positive valence). (*C* and *D*) When the face and body are blurred out, inference of valence and arousal is still possible. Images courtesy of (*Upper Left, Lower Left*, and *Lower Right*) Imgflip and (*Upper Right*) Getty Images/Michael Steele.

Facial Expressions

A facial expression is a facial configuration that conveys a specific, and sometimes unique, meaning. Darwin (5) famously suggested that there are six such unique facial expressions—those used to communicate happiness, surprise, sadness, anger, disgust, and fear. That is, there are six facial configurations that, when observed in isolation, are most of the time interpreted to communicate these emotion categories (6). Indeed, many studies have been completed that confirm this view, although more recent studies have identified a much larger set of expressions, including compound emotions such as happily disgusted as well as other affective signals (2, 7). But context remains essential to correctly interpret facial expressions, as demonstrated in the examples in Figs. 1 and 2 *A* and *B*. In these examples, the emotion category selected by observers when seeing the face alone or in context changes radically.

Some researchers claim that other affect variables like valence (i.e., whether the expression is positive or negative and by how much) and arousal (i.e., the degree of engagement) are what is robustly communicated through facial expressions, not emotion categories (8, 9). However, as we see in Figs. 1 and 2*A*, even the broad concepts of valence and arousal may change when we remove the context from the image.

Although the importance of context in the interpretation of facial expressions has been known for a while, Chen and Whitney (3) now show that context alone may be sufficient to interpret the affective state of a person. In fact, Chen and Whitney show that people are surprisingly good at estimating the affect of individuals, even when their faces and bodies have been blurred out. The image in Fig. 2*D* shows an example. This image corresponds to the one in Fig. 2*A* but has the whole body blurred out, yet the affect concepts of valence and arousal can be accurately and robustly estimated by most naïve observers.

Although it may now seem obvious that context is generally a strong determinant of how a person is feeling, both the fact that context alone is sufficient and the way that different contextual cues are combined to make this inference will come as a surprise to many researchers. First, this means that scientists interested in studying the perception of faces and emotion will need to incorporate the role of scene analysis, objects, and other agents, as well as their interactions in face perception models. In addition, it is reasonable to assume that when we make an inference of how others feel based on context alone, we do so by thinking of how we, ourselves, would feel in that same situation. This suggests that we employ theory of mind to interpret affect in others, not solely face and scene perception mechanisms as is typically assumed. But how all these different cognitive abilities interact to make the final inference is completely unknown. Further research is thus needed to fill in the gaps in our understanding of how we interpret affect in others.

Contextual Information

Facial expressions rarely (if ever) occur in isolation but rather are part of a scene that may include objects and other people. For example, a person smiling while comfortably seated next to others on a beach will be interpreted as expressing joyfulness (i.e., a positive emotion), whereas the same facial configuration in a meeting room with others expressing disbelief may be interpreted as embarrassment (i.e., a negative emotion). It is easy to see that even if the face of the person of interest in both conditions were not visible, we would readily reach the same conclusions. But how do we make such an inference?

It seems clear that when other people are in the scene, their facial expressions and body pose will generally influence our decision. However, is this all of the information observers use to infer the affect of the unseen person? Chen and Whitney's (3) study suggests that this is not the case and that other scene information is also highly relevant.

To identify which low-level (image) and high-level (semantic) features are most relevant, Chen and Whitney (3) designed an extraordinary experiment in which subjects saw videos in four different conditions. The first condition was the original, unaltered video, which is used to define the baseline perception of affect of an individual in the image. Two other conditions blurred either the person we want to infer affect from or the rest of the image (context), which identifies the roles the facial expression and the scene have in our inference, respectively. A final condition blurred the subject and removed the contextual information completely (leaving only a black background), which allows the authors to compute the contribution of biological motion (i.e., how the person moves in the scene, which is still visible when the character is blurred out). It turns out that a simple linear combination of observers' inferred affect in the latter three conditions is almost identical to that given in the unaltered video, although valence was predicted more accurately than arousal. Thus, the actual facial configuration of the subject, the subject's biological motion, and context all appear to play an important role in our interpretation of how a person feels, and, crucially, their contributions appear to be aggregated linearly rather than by means of a more sophisticated cognitive mechanism.

What is not yet known is what specific image cues our visual system uses to estimate affect from bodies, biological motion, or the background scene. It seems obvious to hypothesize that the spatial and temporal relationships among them and the unseen individual play a role. Equally likely is the influence of some key objects, with their categories (e.g., a cake, a knife) as well as their spatial and temporal interactions (e.g., whether the knife is used to cut the cake or threaten someone) having a significant effect. Regarding the background scene (e.g., a beach, a city street, a room), what image or semantic features are most relevant? In addition, one would expect observers' consensus to decrease as context becomes more cluttered or socially complex, but this is not studied in the present work.

Another problem that the current paper leaves unanswered is the significant across-subject variability on the estimates of valence and arousal. As Chen and Whitney (3) report, each subject provides quite a distinct guess on how the individual in the video feels. (Their results are computed over the average of all subjects, as is customary in this type of experiment.) This suggests that other variables, such as personal biases and one's state of mind, are important and will need to be considered in future models (10, 11). These variables may include biases over certain environments (e.g., a video of what appears to be a dangerous neighborhood to one observer but not another), cultural moral beliefs, and personal traumatic experience (whether these have incurred over development or correspond to a recent event) (12).

Another limitation of the present study is that it only evaluated observers' estimates of valence and arousal from a small number of videos. How well and, most importantly, under which conditions can we correctly estimate valence, arousal, and other affective variables from context alone in video sequences and still images?

Significance of These Results

Computer algorithms capable of imitating our face-recognition abilities are becoming ubiquitous, yet almost none of them use the distinct visual cues identified by Chen and Whitney (3) in PNAS, but rather are limited to the analysis of faces in isolation. It is now clearer than ever that contextual information must be included if we expect these computer vision systems to make good inferences of the affect of the people they analyze. Moreover, and as mentioned earlier, people most likely employ theory of mind to estimate how others feel, but artificial intelligence algorithms have yet to imitate this high-level cognitive ability.

In addition, limitation in reading facial expressions of emotion has been listed as an important symptom in several psychopathologies; for example, autism spectrum disorder and post-traumatic stress disorder (13). When evaluating these individuals, however, many researchers, including members in my research group, have not found a major disruption of the recognition of affect from facial expressions seen in isolation. The results of Chen and Whitney (3) suggest that we should now consider other atypical visual analyses, especially of scenes, objects, and other people. If no limitation in the analysis of scenes and objects were found, as previous research suggests, it might be indicative of a limitation in the communication between distinct cognitive processes, a difficulty in creating affective gestalts, deficits in theory of mind, or a predictive problem (14). Further research is needed to evaluate these hypotheses.

Finally, and crucially, we still do not know which brain regions of interest (ROIs) are involved in the recognition of affective context and how these communicate with already known areas. ROIs for the recognition of facial muscle articulations and biological motion (2) as well as affective variables (15) have been identified. It is logical to assume that ROIs involved in visual analysis of scenes, objects, and bodies are part of the visual interpretation of context. But does the brain employ other mechanisms to perform this inference? And, how do all these ROIs interact with one another to create the conscious affective percept we all experience? Given the fact that a simple linear combination of each visual cue identified by Chen and Whitney (3) was shown to explain people's affective inference, we might expect to find a straightforward interaction between these ROIs in subsequent studies. I, for one, cannot wait to find out how the brain solves such a complex problem quickly and efficiently while making it look effortless.

1 Phillips PJ, et al. (2018) Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proc Natl Acad Sci USA* 115:6171–6176.

2 Barrett LF, Adolphs R, Marsella S, Martinez AM, Pollak S, Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychol Sci Public Interest*, in press.

3 Chen Z, Whitney D (2019) Tracking the affective state of unseen persons. *Proc Natl Acad Sci USA*, 10.1073/pnas.1812250116.

4 Duchenne de Boulogne G-B (1990) *The Mechanism of Human Facial Expression* (Cambridge Univ Press, New York).

5 Darwin C (2009) *The Expression of the Emotions in Man and Animals* (Penguin Classics, London).

6 Ekman P (2016) What scientists who study emotion agree about. *Perspect Psychol Sci* 11:31–34.

7 Keltner D, Sauter D, Tracy J, Cowen A (February 7, 2019) Emotional expression: Advances in basic emotion theory. *J Nonverbal Behav*, 10.1007/s10919-019-00293-3.

8 Chen C, et al. (2018) Distinct facial expressions represent pain and pleasure across cultures. *Proc Natl Acad Sci USA* 115:E10013–E10021.

9 Russell JA (2003) Core affect and the psychological construction of emotion. *Psychol Rev* 110:145–172.

10 Barrett LF (2017) *How Emotions Are Made: The Secret Life of the Brain* (Houghton Mifflin Harcourt, Boston).

11 Barrett HC, et al. (2016) Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proc Natl Acad Sci USA* 113:4688–4693.

12 Pollak SD, Kistler DJ (2002) Early experience is associated with the development of categorical representations for facial expressions of emotion. *Proc Natl Acad Sci USA* 99:9072–9076.

13 American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* (American Psychiatric Publishing, Washington, DC), 5th Ed.

14 Sinha P, et al. (2014) Autism as a disorder of prediction. *Proc Natl Acad Sci USA* 111:15220–15225.

15 Wang S, et al. (2014) Neurons in the human amygdala selective for perceived emotion. *Proc Natl Acad Sci USA* 111:E3110–E3119.