

Semantic Space Theory: Data-Driven Insights Into Basic Emotions

Dacher Keltner¹, Jeffrey A. Brooks, and Alan Cowen²

Department of Psychology, University of California, Berkeley, and Hume AI, New York, NY

Current Directions in Psychological Science
1–8

© The Author(s) 2023

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09637214221150511

www.psychologicalscience.org/CDPS



Abstract

Here we present semantic space theory and the data-driven methods it entails. Across the largest studies to date of emotion-related experience, expression, and physiology, we find that emotion is high dimensional, defined by blends of upward of 20 distinct kinds of emotions, and not reducible to low-dimensional structures and conceptual processes as assumed by constructivist accounts. Specific emotions are not separated by sharp boundaries, contrary to basic emotion theory, and include states that often blend. Emotion concepts such as “anger” are primary in the unfolding of emotional experience and emotion recognition, more so than core affect processes of valence and arousal. We conclude by outlining studies showing how these data-driven discoveries are a basis of machine-learning models that are serving larger-scale, more diverse studies of naturalistic emotional behavior.

Keywords

emotion, expression, semantic space, expression models

Emotions are complex states involving appraisals, expressions, physiological reactions, and patterns of feeling and thought (Keltner, Oatley, & Jenkins, 2018). Central to their understanding is the study of how emotional behaviors are associated with social contexts, neurophysiological responses, and self-reported experiences or perceptions (Scherer, 2005). At stake in such research are many questions: What are the emotions? Are specific emotions separated by clear boundaries, or discrete? What is primary in emotional experience, appraisals of valence and arousal or distinct emotions?

Here we present a new approach to these questions offered within semantic space theory. We first outline this theory and its central tenet that the latent dimensions of emotion can be derived quantitatively from data (Cowen, Sauter, et al., 2019). We then survey how recent studies of emotion-related experience, expression, and physiology offer tests of basic emotion theory and constructivist approaches. We conclude by considering advances in machine-learning methods for measuring emotion-related behavior.

Semantic Space Theory

Central to emotion science is the reliance upon self-reports or measures of behavior to understand how appraisals of stimuli give rise to varied emotional

experiences. To approach such complexity, semantic space theory posits that emotion-related behaviors reside in what we call a semantic space (Cowen & Keltner, 2018). Much as other subjective phenomena—colors, tastes, physical sensations—are organized within broader representational spaces, the same is true of emotion. The emotional experiences or expressions that arise in an interaction with friends, for example, are determined by dynamic trajectories within a semantic space that capture how eliciting stimuli, experiences, and expressive behaviors relate to one another. Three properties define such semantic spaces and their variation across individuals and cultures (Cowen & Keltner, 2018, 2021).

A first is their *dimensionality*, or number of distinct dimensions or kinds of emotion that organize a semantic space. Dimensions are not to be confused with discrete categories or clusters, which are terms that describe the distribution of emotion (see next paragraph). Dimensions are typically continuous features that can combine in any number of ways and are uncovered in complex data with multidimensional reliability

Corresponding Author:

Dacher Keltner, Department of Psychology, University of California, Berkeley

Email: keltner@berkeley.edu

Table 1. Basic Emotion Theory and Constructivist Hypotheses About Semantic Spaces of Emotion

Theory	Dimensionality	Distribution	Conceptualization
Basic emotion theory	Emotion-related phenomena involve five to 10 qualitatively distinct underlying states	The boundaries between emotional states are discrete	Specific emotions align with distinct dimensions that are primary in experience and recognition
Constructivist theory	Emotion-related phenomena involve two or three underlying dimensions of neurophysiological response	The boundaries between emotional states are not discrete	The dimensions are valence and arousal, and emotions are language-dependent constructions along these dimensions

analyses (see next paragraph). In principle, a single emotion, such as “fear,” can contribute to multiple distinct dimensions; we find, though, that dimensions of semantic spaces of emotion most typically correspond to individual emotions.

The second property of a semantic space is the *distribution* of states within the space—that is, the geometric arrangement of emotions and boundaries between them. Are different kinds of emotion separated by distinct boundaries and organized within clusters, as presupposed by the phrase “discrete categories of emotion”?

Whether emotion-related phenomena are better described by emotion categories or broader appraisals is a question regarding *conceptualization*. Is there more explanatory value in the use of specific emotion concepts (e.g., “sympathy,” “anger”) or more general psychological features, most notably of valence and arousal, to explain emotion-related experience, expression, or physiology?

The methods typically used in emotion science assume that the dimensionality, distribution, and conceptualization of emotion are known in advance, and preclude uncovering these properties from the data. A typical study might focus on four to six emotions or two dimensions presumed to explain emotion-related responses, typically evoked using a limited number of stimuli and characterized with self-report measures that capture four to 10 emotions (Cowen & Keltner, 2017). The dimensionality of such data will, by definition, be low. The distribution of states will be biased toward isolated, discrete clusters. Forced-choice measures that focus on a narrow array of self-reports do not allow for the modeling of how people conceptualize emotion.

Given these concerns, we developed data-driven methods to map a semantic space of any emotion-related response. Participants are presented with a vast array of stimuli—short films, facial expressions, samples of music—instead of a limited number of prototypical stimuli. Participants report on 50 emotions and appraisals (Cowen & Keltner, 2017). Finally, with parallels to new

approaches to brain imaging data (e.g., Davis et al., 2014), new statistical techniques reveal the patterns of relations between emotions (see Cowen & Keltner, 2017, 2018). Traditional methods, such as correlations, *t* tests, and emotion recognition accuracy measures, introduce assumptions about the structure of emotion and focus on testing one-to-one mappings between specific emotions and stimuli. The methods we have developed, by contrast, assess the patterns of covariation between self-reports of emotion-related experience, expression, and physiology. Whereas traditional dimensionality reduction techniques, such as factor-analytic and principal components analyses, use heuristic criteria, such as scree tests, to reduce the dimensionality of data sets, we have focused on how the multidimensional factors maximally covary across multiple data sets (such as emotion judgments from two cultures), which allows a characterization of the latent structure of multidimensional data (see Cowen et al., 2021; Cowen, Elfenbein, et al., 2019; Cowen, Sauter, et al., 2019; Demszky et al., 2020).

With these data-driven methods, we have carried out studies testing the claims of basic emotion theory and constructivist accounts, which we present in Table 1 (see Cowen & Keltner, 2020, 2021). Traditionally, these two approaches diverge in their accounts of the dimensionality of emotion, with basic emotion theorists claiming that emotion-related responses involve a small number of distinct states and constructivist accounts representing emotion along two or three dimensions of “core affect” (typically conceptualized as valence and arousal). A traditional claim of basic emotion theory is that the boundaries between emotions are discrete, which pertains to the distribution of emotions. Long debated, but little tested, are contrasting claims about conceptualization: In basic emotion theory, specific emotions, such as “awe” and “disgust,” describe distinct dimensions of emotion-related phenomena, whereas in constructivist approaches, emotions are constructed meanings within the two- or three-dimensional space of core affect. Our recent studies, spanning thousands

(Cordaro et al., 2016), prosody (Cowen, Laukka, et al., 2019), and full-body expression (Monroy et al., 2022), and cross-cultural similarities (between 50% and 70%) in facial-bodily expression of emotion (Cordaro et al., 2018, 2020) and mimicry of others' vocal expression (Brooks et al., 2022). Across semantic spaces of experience, expression, and physiology, then, 20 to 25 distinct kinds of emotion are found.

Specific Emotions Are Heterogenous and Not Discrete

In a typical study in emotion science, prototypical stimuli are sorted into a handful of categories, and elicited responses are judged in terms of their correspondence to an expected response (Cowen, Sauter, et al., 2019). Mismatches are treated as challenges to the idea that emotion categories are stable and coherent (e.g., Barrett et al., 2019). The methods we have outlined here allow the structure of emotion to emerge from the data itself. No assumptions are made about the discreteness of emotion.

As evident in Figure 1, emotion blends are common. Within a category of emotion there are more and less prototypical experiences and expressions (Ekman, 1992). The frequency of emotion blends highlights the need to model the heterogeneity of emotion-related responses within a category, rather than treating it as noise, inaccuracy, or category incoherence; data-driven methods allow this.

A central tenet of basic emotion theory is that there are discrete boundaries between emotions (e.g., Ekman, 1992). This we find to be true for select emotions (see desire and craving in Fig. 1a). More typically, the boundaries between emotions are fuzzy, not discrete. In addition, certain cognitive appraisals arrayed on dimensions bridge distinct emotions. Appraisals of effort, for example, account for transitions from experiences of disgust to anger (see Cowen & Keltner, 2017). Categories of emotion, then, are heterogenous and not separated by discrete boundaries and are bridged by gradients of meaning.

Specific Emotions Are Primary in Emotion Conceptualization

What structures how people conceptualize an emotion-related response, specific emotions, or appraisals of valence and arousal? Answering this question has proven elusive because of the empirical focus on a limited array of emotions elicited with equally impoverished stimuli and self-report terms.

Three kinds of evidence from our studies reveal specific emotions to be primary in the conceptualization of emotion. First, across studies, valence and arousal explain only a small fraction—typically around 25%—of

the variance in experiences of states like “disgust” or “awe”; reported categories of emotion, by contrast, capture nearly the entirety of the variance in valence and arousal. Valence and arousal would appear to be retrospective inferences about emotional phenomena that are fully explained by, but do little to explain, more primary, stimulus-specific representational processes.

The cross-cultural stability of self-reports offers a second test of the conceptualization primacy question. In cross-cultural research, studies typically compare the similarity with which individual members of two cultures label a stimulus with the same terms, which conflates individual and cross-cultural variation. In such studies, the maximum level of similarity that can be observed across cultures is the product of the interrater reliability within each culture. To control for individual-level variation, we use techniques that compare the stability across cultures with which specific emotions and dimensions of valence and arousal are associated with expressions and elicitors of emotion. For example, Figure 2 presents the results from a study of the recognition of emotion from 2,519 speech samples from five cultures (Cowen, Laukka, et al., 2019). Correlations between Indian and U.S. participants' ratings reveal greater cross-cultural stability in how the voice conveys specific emotions than valence and arousal. Similar results were observed in Chinese and U.S. participants' ratings of over 1,500 samples of music (Cowen et al., 2020).

Recent neuroscientific studies yield a third kind of evidence pointing to the primacy of specific emotions in conceptualization processes. In an analysis of the temporal unfolding of the brain's representations of emotion vocalizations in terms of self-reports, specific emotions manifested earlier in the brain's response and valence and arousal later—perhaps as more “abstract,” even epiphenomenal, forms of conceptualization, posing problems for constructivist accounts that posit that valence and arousal are primary in the conceptualization of emotion (Giordano et al., 2021). In the functional magnetic resonance imaging study by Horikawa and colleagues (2020) referred to earlier, self-reports of dozens of emotions explained patterns of evoked brain activity better than ratings of valence and arousal in every brain region measured, both cortical and subcortical, suggesting that even in subcortical emotion-related processing, specific emotions are primary.

A Synthesis and Future of Emotion Science

In studies spanning multiple cultures, thousands of stimuli, and large and diverse samples, we find that 21 emotions meet Ekman's criteria for “basic” emotions, being associated with distinct antecedents, experiences, expressions, and neurophysiological correlates (Ekman,

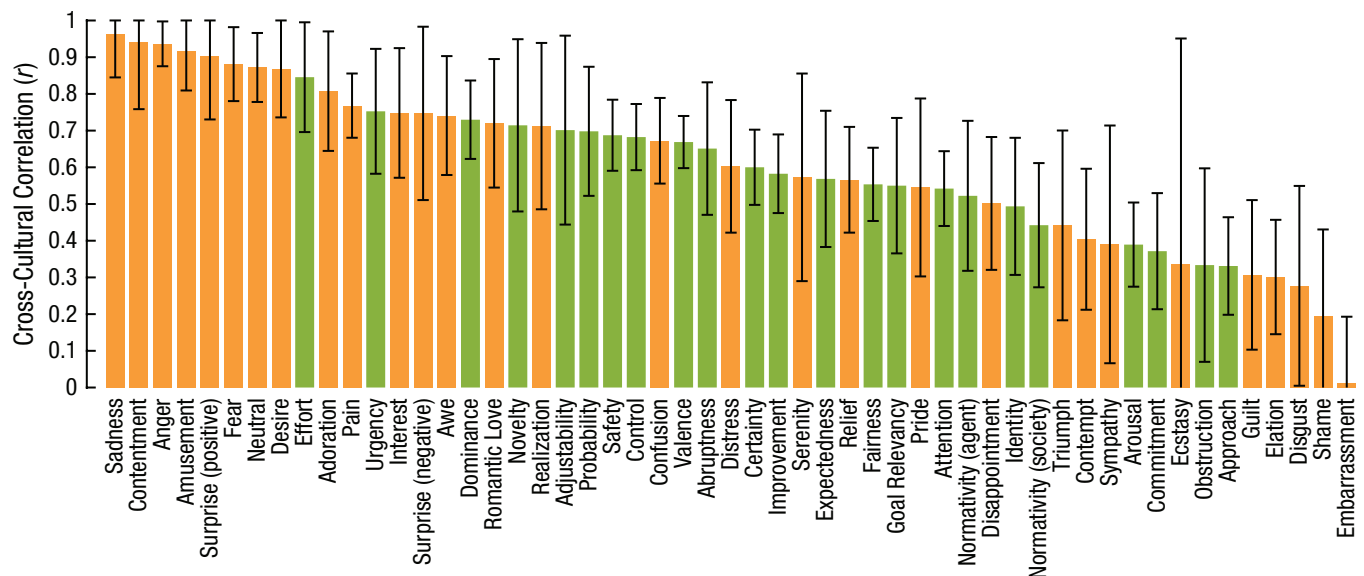


Fig. 2. Correlations across Indian and U.S. participants' ratings of emotional prosody. Error bars represent standard error (estimated by bootstrapping across participants).

1992; see Cowen & Keltner, 2021). These emotions include 10 “negative” states, including anger, anxiety, boredom, confusion, disgust, embarrassment, fear, pain, sadness, and shame, and 10 “positive” states, including amusement, awe, contentment, desire, excitement, love, interest, pride, sympathy, and triumph (see also Weidman & Tracy, 2020), as well as surprise.

Our findings are descriptive. We have not explained how these emotions arise, unfold, and vary across cultures—a task for componential, discrete, and dimensional appraisal approaches, perhaps aided, we hope, by the data-driven methods summarized here. Nor have we offered functional accounts of why such a rich space of emotions emerged in hominid evolution, although findings dovetail with recent explanations of the self-conscious emotions (Sznycer & Cohen, 2021), hierarchically relevant emotions (Tracy et al., 2020), attachment- and knowledge-related emotions (Shiota et al., 2017), and aesthetic emotions (Juslin, 2019; Schindler et al., 2017).

Emotion science has been constrained by its reliance on small and homogenous samples and time-consuming approaches to behavioral measurement. Grounded in our data-driven discoveries, we have developed machine-learning “expression models” based on these methodological principles: (a) build large-scale data sets of naturalistic samples of emotional behavior, accompanied by psychologically valid measures of the meaning of the context and the behavior (as ascribed by the people forming the behavior and outside perceivers), and (b) train machine-learning models to predict the context or meaning of behaviors directly from raw auditory or visual recordings. We have used these methods to train machine-learning models to output

objective measures of facial and vocal expressions based on hundreds of thousands of emotionally rich images, videos, and audio recordings.

In one study, annotations of hundreds of thousands of videos were used to train a machine-learning model that could accurately predict judgments of 16 kinds of facial expression. Importantly, the model was exposed only to pixels on the face itself, and normalization methods were used to remove effects of viewpoint and lighting. We then applied the model to a separate set of six million videos from around the world, each annotated using both video- and text-based machine-learning models to determine its social context (Cowen et al., 2021). Across 144 countries, 16 facial expressions reliably arose in naturalistic social contexts, and at minimum, there was a 70% overlap between any two broad cultural regions (e.g., western Europe, East Asia) in the covariation between facial expressions and specific social contexts.

With large-scale, experimentally controlled data sets that eliminate correlations between emotional behaviors and the traits, clothing, or environment of the people forming them, we have developed models for measuring facial expressions, vocal bursts (Brooks et al., 2022), text, and prosody (see <https://hume.ai/> for more information). In two such experiments, we characterized the semantic spaces of how thousands of participants from Asian, African, South American, and western European countries mimic naturalistic facial and vocal expressions (Brooks et al., 2022). Models trained on these diverse data sets (involving hundreds of thousands of mimicked expressions) uncovered considerable cross-cultural similarity in 28 kinds of facial

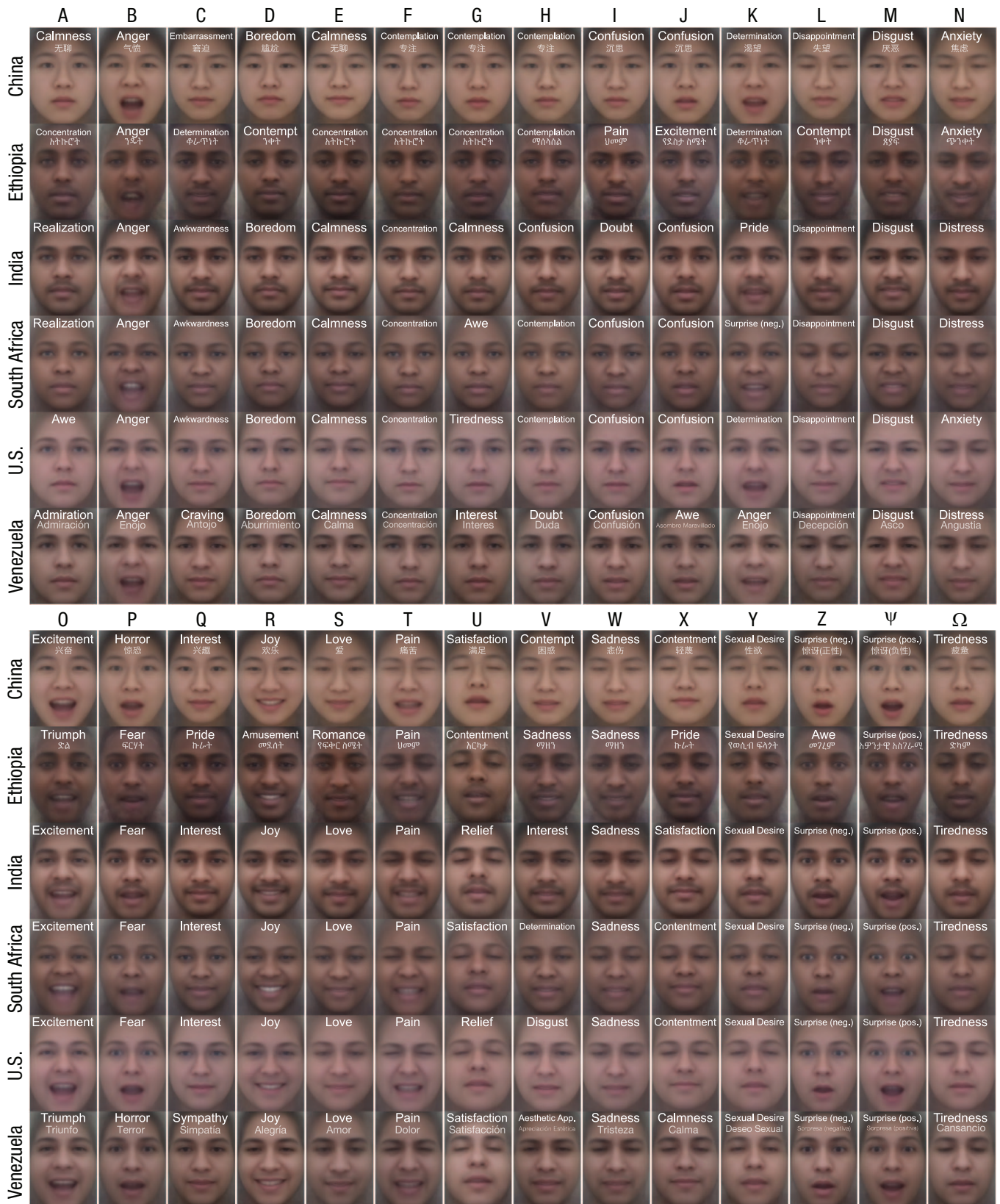


Fig. 3. Twenty-eight dimensions of facial expression in six countries discovered by an expression model (Brooks et al., 2022). We used a deep neural network (DNN) to find dimensions of facial expression across six countries, independent of facial appearance and context, by averaging ratings of 4,659 images in each country and tasking the DNN with predicting these country-wise average ratings solely from imitations of each expression. We present morphs of the average facial movements associated with the 28 dimensions and the concept most reliably associated with each dimension.

expression (see Fig. 3), overlapping 63% in meaning across six countries, and 24 kinds of distinct vocal expression, overlapping 79% in meaning across five countries.

Semantic space theory and new data-driven methods allow for a new path forward in emotion science for finding answers to perennial questions about emotion in larger-scale, more diverse studies of naturalistic behavior.

Recommended Reading

- Cowen, A. S., & Keltner, D. (2017). (See References). The first application of semantic space theory and its data-driven methods, finding that 27 distinct emotions are felt in response to highly evocative GIFs.
- Cowen, A. S., & Keltner, D. (2021). (See References). Offers a more detailed study of the methods we profile in this article and a broader survey of the neurophysiology of distinct emotions.
- Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2021). (See References). The first article to extend semantic space theory to machine-learning approaches to emotional expression, finding that naturalistic facial expressions of 16 emotions covary with specific contexts in highly similar ways across 144 countries.
- Keltner, D., Oatley, K., & Jenkins, J. M. (2018). (See References). Provides a broad overview of what has been learned in emotion science.

Transparency

Action Editor: Robert L. Goldstone


Editor: Robert L. Goldstone

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

ORCID iDs

Dacher Keltner  <https://orcid.org/0000-0001-9061-5292>

Alan Cowen  <https://orcid.org/0000-0002-8381-5883>

References

- Asutay, E., Genevsky, A., Barrett, L. F., Hamilton, J. P., Slovic, P., & Västfjäll, D. (2021). Affective calculus: The construction of affect through information integration over time. *Emotion, 21*(1), 159–174.
- Bakker, I., van der Voordt, T., Vink, P., & de Boon, J. (2014). Pleasure, arousal, dominance: Mehrabian and Russell revisited. *Current Psychology, 33*(3), 405–421. <https://doi.org/10.1007/s12144-014-9219-4>
- Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(1), 1–23.
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest: A Journal of the American Psychological Society, 20*(1), 1–68.
- Brooks, J. A., Kim, L., Opara, M., Keltner, D., Fang, X., Monroy, M., Corona, R., Tzirakis, P., Baird, A., Metrick, J., Tadesse, N., Zegeye, K., & Cowen, A. S. (2022). Deep learning reveals what facial expressions mean to people in different cultures [Manuscript in preparation].
- Brooks, J. A., Tzirakis, P., Baird, A., Kim, L., Opara, M., Fang, X., Keltner, D., Monroy, M., Corona, R., Metrick, J., & Cowen, A. S. (2022). Deep learning reveals what vocal bursts express in different cultures. *Nature Human Behaviour, 7*(2), 240–250. <https://doi.org/10.1038/s41562-021-01073-0>
- Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion, 17*, 117–128.
- Cordaro, D. T., Sun, R., Kamble, S., Hodder, N., Monroy, M., Cowen, A., Bai, Y., & Keltner, D. (2020). The recognition of 18 facial-bodily expressions across nine cultures. *Emotion, 20*(7), 1292–1300.
- Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion, 18*(1), 75–93.
- Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2019). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist, 74*(6), 698–712.
- Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2020). What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. *Proceedings of the National Academy of Sciences, USA, 117*(4), 1924–1934.
- 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, USA, 114*(38), E7900–E7909.
- Cowen, A. S., & Keltner, D. (2018). Clarifying the conceptualization, dimensionality, and structure of emotion: Response to Barrett and colleagues. *Trends in Cognitive Sciences, 22*(4), 274–276.
- Cowen, A. S., & Keltner, D. (2020). What the face displays: Mapping 28 emotions conveyed by naturalistic expression. *American Psychologist, 75*(3), 349–364.
- Cowen, A. S., & Keltner, D. (2021). Semantic space theory: A computational approach to emotion. *Trends in Cognitive Sciences, 25*(2), 124–136.
- Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2021). Sixteen facial expressions occur in similar contexts worldwide. *Nature, 589*(7841), 251–257.
- Cowen, A. S., Laukka, P., Elfenbein, H. A., Liu, R., & Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature Human Behaviour, 3*(4), 369–382.
- Cowen, A. S., Sauter, D., Tracy, J. L., & Keltner, D. (2019). Mapping the passions: Toward a high-dimensional

- taxonomy of emotional experience and expression. *Psychological Science in the Public Interest*, 20(1), 69–90.
- Davis, T., LaRocque, K. F., Mumford, J. A., Norman, K. A., Wagner, A. D., & Poldrack, R. A. (2014). What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *NeuroImage*, 97, 271–283.
- Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). *GoEmotions: A dataset of fine-grained emotions*. <http://arxiv.org/abs/2005.00547>
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169–200.
- Giordano, B. L., Whiting, C., Kriegeskorte, N., Kotz, S. A., Gross, J., & Belin, P. (2021). The representational dynamics of perceived voice emotions evolve from categories to dimensions. *Nature Human Behaviour*, 5(9), 1203–1213. <https://doi.org/10.1038/s41562-021-01073-0>
- Horikawa, T., Cowen, A. S., Keltner, D., & Kamitani, Y. (2020). The neural representation of visually evoked emotion is high-dimensional, categorical, and distributed across transmodal brain regions. *iScience*, 23(5), 101060.
- Juslin, P. N. (2019, May 23). *Musical emotions explained: Unlocking the secrets of musical affect*. Oxford Academic. <https://doi.org/10.1093/oso/>
- Keltner, D., Oatley, K., & Jenkins, J. M. (2018). *Understanding emotions*. Wiley Global Education.
- Monroy, M., Cowen, A. S., & Keltner, D. (2022). Intersectionality in emotion signaling and recognition: The influence of gender, ethnicity, and social class. *Emotion*, 22(8), 1980–1988.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1), 145–172.
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729.
- Schindler, I., Hosoya, G., Menninghaus, W., Beermann, U., Wagner, V., Eid, M., & Scherer, K. R. (2017). Measuring aesthetic emotions: A review of the literature and a new assessment tool. *PLOS ONE*, 12(6), Article e0178899. <https://doi.org/10.1371/journal.pone.0178899>
- Shiota, M. N., Campos, B., Oveis, C., Hertenstein, M. J., Simon-Thomas, E., & Keltner, D. (2017). Beyond happiness: Building a science of discrete positive emotions. *American Psychologist*, 72(7), 617–643.
- Sznycer, D., & Cohen, A. S. (2021). How pride works. *Evolutionary Human Sciences*, 3, E10. <https://doi.org/10.1017/ehs.2021.6>
- Tracy, J. L., Mercadante, Z., Witkower, Z., & Cheng, J. T. (2020). The evolution of pride and social hierarchy. *Advances in Experimental Social Psychology*, 62, 51–114.
- Weidman, A. C., & Tracy, J. L. (2020). A provisional taxonomy of subjectively experienced positive emotions. *Affective Science*, 1(2), 57–86.