Debate: Keltner and Cordaro vs. Fridlund vs. Russell

Dacher Keltner is a full professor at UC Berkeley and director of the Berkeley Social Interaction Lab (http://socrates.berkeley.edu/~keltner/) and faculty director of the Greater Good Science Center (http://greatergood.berkeley.edu). Dacher's research focuses on the biological and evolutionary origins of compassion, awe, love, and beauty, emotional expression, and power, social class, and inequality. Dacher is the co-author of two textbooks, as well as the best-selling *Born to Be Good: The Science of a Meaningful Life*, and *The Compassionate Instinct*. Dacher has published over 190 scientific articles, he has written for the *New York Times Magazine*, *The Wall Street Journal*, *The London Times*, and *Utne Reader*, and has received numerous national prizes and grants for his research. He served as a consultant for Pixar’s *Inside Out*, and WIRED magazine recently rated Dacher’s podcasts from his course Emotion as one of the five best educational downloads, and the *Utne Reader* selected Dacher for one of its 50 2008 visionaries.

Daniel T. Cordaro received his Ph.D. from the University of California, Berkeley where he directed the Universal Expression Project – a program that aimed to decode the universal language of human expression across eleven cultures. The Universal Expression Project found evidence in both industrialized and remote cultures for not only universal patterns in emotional expression, but also some fascinating cultural variations and individual differences. Daniel is currently a postdoctoral research associate at the Yale Center for Emotional Intelligence, where he is exploring new lines of research on contentment and creating interventions for adults and children on cultivating emotional wellbeing in everyday life. He is also developing a new test of emotional intelligence to assess five skills: recognizing, understanding, labeling, expressing, and regulating emotions. This work aims to investigate new ways of assessing emotional skills and measure how emotional intelligence plays a role in human wellbeing.

James A. Russell is professor of psychology at Boston College. He has published over one hundred scientific papers, all on some aspect of emotion. His current work focuses on an approach to emotion called psychological construction, especially on ways to integrate this approach with other research programs such as appraisal theory and social construction. James was Editor-in-chief of Emotion Review from 2007 – 2014. He has co-edited *The Psychological Construction of Emotion* (2014) with Lisa Feldman Barrett, and *The Psychology of Facial Expression* (1997) with Jose Fernandez-Dols.

Alan Fridlund is Associate Professor in the Department of Psychological and Brain Sciences, University of California, Santa Barbara, and a member of its faculty in Developmental and Evolutionary Psychology. His work
has concerned the history of psychology, emotion and nonverbal behavior, human evolution and Darwiniana, and morality and punishment. He co-authored three editions of *Gleitman’s Psychology* (Norton), authored *Human Facial Expression: An Evolutionary View* (Academic Press, 1994), and has written an assortment of more focused papers. Though trained fundamentally in animal behavior and physiological psychology, he has maintained an active clinical psychology practice for 30 years, focusing on people with serious mental disorders and co-morbid medical illness.

The Great Expressions Debate

August 2015 – Andrea Scarantino (AS): Thank you very much for your highly informative manifesto papers, which provide an up-to-date overview of three influential traditions in the study of facial expressions: Basic Emotion Theory (manifesto here), the Behavioral Ecology View (manifesto here) and Psychological Constructionism (manifesto here). And thank you for agreeing to participate in this Q&A, in the course of which we will explore some of the central issues raised in the three pieces, hopefully unveil new areas of agreement and get clear on the nature and scope of the remaining disagreements. I begin with two clarificatory questions for each manifesto piece, and then formulate three open questions for all debaters.

Let me start from Keltner and Cordaro’s manifesto piece. Dacher and Daniel, a significant novelty in your piece is the rejection of BET’s standard assumption that emotional expressions must be momentary and facial. On the contrary, you argue that Darwin himself was clear that emotional expressions are extended in time (they convey information about what emotion one is undergoing as they dynamically unfold) and multimodal (they do not just involve the face, but also the voice and the body).

Is this proposal meant to replace Ekman’s suggestion that for each basic emotion family there is a “facial theme” (with some variations) that is distinctive of it? For example, Ekman told us that “in all members of the anger family the brows are lowered and drawn together, the upper eyelid is raised and the muscle in the lips is tightened”. Is your suggestion that there is no facial configuration distinctive of the anger, fear, disgust, surprise, happiness, and sadness families, as there is no facial configuration distinctive of embarrassment?

Keltner and Cordaro (K-C): In many ways BET was founded on the early discoveries of Ekman, Izard, and others focusing on facial muscle movements, discoveries that were enabled by specific methodologies of the times, most notably anatomically based facial action coding and still photography and video analysis. But since those
discoveries, the study of emotional signaling, as we review, has advanced considerably, and it now considers multiple modalities – face, gaze, voice, head movements, bodily movements, tactile contact — and the dynamic unfolding of expressions. This has allowed several emotions neglected in the early phases of expression science back to center stage. In light of this robust new evidence, what facets of BET are in need of revision?

First, we think it is clear that there are emotions that do not have distinct facial expressions. This in itself represents an important departure from the strictest early formulations of BET, which focused intensively upon facial muscle movement as a defining criterion of emotion, but aligns with later versions, for example that offered by Ekman in 1992, that emotions must have distinctive signals, which include other modalities than the face, including vocalizations.

For example, awe has a clear vocal pattern, readily recognized in emotion recognition studies of participants from different parts of the world (Cordaro et al., in press), but we think, in light of select studies, that its facial display of inner eyebrows pulled up and in and gaze directed upward will not prove to be a reliable signal on its own. Or take the case of gratitude. Gratitude is featured in different classifications of emotion, and has many of the subjective qualities attributed to more well studied emotions – it is brief, of quick onset, and triggered by clear antecedents, namely the sense that one has received a gift from another person or people. And while little at this stage is known about its physiology, gratitude does have a signal, one with clear antecedents in the behaviors of others species, but it is not in the face.

Specifically, as we review in our essay, humans can communicate gratitude with brief patterns of tactile contact that resemble in form, context, and function the grooming behavior of nonhuman primates. Importantly, still other studies find that gratitude cannot be communicated with any distinct facial expression, nor with any vocal display (e.g., Simon-Thomas et al., 2009). This kind of evidence suggests that facial expression is not the *sine qua non* of emotion, and is a significant reason why our reformulated BET focuses on multimodal, dynamic displays of emotion.

There are other emotions that do involve patterns of facial muscle movements, but that require other expressive behaviors — bodily movements, gaze activity, head movements – to signal the distinct emotion. This is most clearly true of the following emotions: pride – which involves a smile but requires head movements back and postural expansion to signal the state; embarrassment, which involves a controlled smile but requires head movements down and gaze aversion to signal the state; and love, which also involves a smile, but requires extended eye contact, head tilts sideways, and open handed gestures to signal the state (Gonzaga et al., 2001; Keltner, 1995; Tracy & Robins, 2004).

Here again we depart importantly from earlier formulations of BET, in particular that of Ekman, which did not consider these states to be emotions. Yet in following the methodologies laid out in these early formulations of BET, researchers are finding it to be increasingly clear that these three states – pride, embarrassment, and love, as well as others – meet the criteria laid out in BET to be considered distinct emotions. Had the field only focused on facial muscle movements, we would have learned little about the displays of these three emotions. By looking to other expressive modalities, signals have been uncovered, and emotion taxonomies have evolved.

This, then, represents an important expansion upon BET – there are proving to be a significant number of emotions – embarrassment, love, pride, awe, gratitude for example – that do not have distinct facial expressions, but that share other features of more well studied emotions. Past conceptualizations have treated these states as more extended affective phenomena that are not basic emotions – (love is a “plot”; see Ekman, 1992), or they have not
considered these states (pride), or they have been agnostic as to whether these states will eventually demonstrate the emotion-like characteristics (e.g., awe; Ekman, 1992). As the evidence guided by BET mounts, we believe these states will meet the criteria for being considered emotions, as will other states (for full set of possibilities, see Keltner & Lerner, 2010).

Given the preceding arguments, can one infer that there are no distinct facial muscle movements, as postulated by BET, that covary with emotions like anger, fear, or disgust? The answer is a qualified no. There are reviews showing that self-reports of experiences of emotions like anger and fear do correlate with the predicted facial muscle movements from BET (Matsumoto et al., 2008). But, the critique of Russell in this debate and elsewhere points to necessary modifications. First, the relationship between experience and muscle movement is weak: correlations between emotion-specific muscle movements – what Ekman has called “reliable” muscles – and self-reports of the experience of emotion tend to range from .25 to .40. Second, the prototypical patterns are less frequent than assumed, although we hasten to note that the kinds of situations studied in the lab do not resemble the prototypical elicitors of emotions, in particular when it comes to emotions like fear or anger.

This leads us to conclude that the emotions of intense focus early in BET – anger, disgust, fear, happiness, sadness, and surprise – do have distinct facial muscle movements, but the patterns of those movements vary according to the individual and context, and they are only weakly associated with the experience of emotion. We think the variability in these expressions represents a fascinating area of unexplored variance in the field, and is likely to be illuminated by studying the influence of temperament, culture, and context upon facial expression. We think the weak associations between facial muscle movement and experience not only point to methodological problems – e.g., retrospective reports of emotion that are increasingly biased as the time between emotion eliciting stimulus and report increases – but necessary conceptual modifications of BET, most notably a systematic focus on context and intention in crafting the nature of the expression, as long suggested by Fridlund.

AS: Thanks Dacher and Daniel, let me briefly follow up. I was struck by the fact that in your answer you never once used the term basic, and included emotions like love that were explicitly excluded from the set of basic emotions by Ekman, and handed out to a theory of emotional plots. Is your view at this point that the qualifier “basic” has lost its function, in that all emotions are basic on your expanded notion? Or do you still acknowledge the existence of non-basic emotions, and if so how are they related, if at all, to basic emotions?

K-C: Early formulations of BET reserved the term “basic” for a limited set of emotions, those that are sufficiently distinct in different response systems from other emotions, and that can be understood as serving specific functions within an evolutionary account of emotion. The use of this term was useful for the times – the 1980s and 1990s — in making the strong case that emotions are distinct, in particular with respect to physiological patterning, and in beginning to make the argument that emotions can be thought of as adaptations, shaped by our evolution, a claim that has been significantly elaborated upon since early formulations of BET.

Nearly 25 years later, we think that the use of the term “basic” poses more problems than the benefits it introduces. In the most general sense, we believe all emotions have distinct profiles and serve some evolutionary function; so the term “basic” to refer to some emotions but not others is unnecessary. Just as critically, the term “basic” separates emotions into those that are “basic” and those that are “not basic.” It is not clear at all what a non-basic emotion might be. Less evolved? Less distinct? Less physiological? Less clearly observed in other species? It’s even harder to imagine what kind of data would sort different affective states into the categories of “basic” emotion and “non basic” emotion, or who, ultimately, would make the designation. For these and other reasons, we do not
think it is productive to refer to some emotions as basic, and instead prefer terms like “well studied.”

We additionally note that some of the concepts used to refer to emotions not meeting the basic emotion criteria in past iterations of BET, such as “emotion plot” to describe love, are also problematic, and do not fare well when stacked up against empirical data. For example, recent studies find that brief experiences of love have distinct patterns of signaling behavior, specific antecedents, and even specific physiological correlates — suggesting that love, like emotions such as anger or fear, meets central criteria in the BET tradition (Gonzaga et al., 2001, 2006).

Ultimately, researchers guided by BET are interested in how emotions differ from one another, and how those distinct emotions shape our inner and outer worlds. They vary in their commitments to evolutionary approaches, and focus on different response systems, but are united by the question of figuring out the forms and functions of different emotions. Since early formulations of BET and its focus on a limited set of emotions, the purview of BET has expanded dramatically, focusing on a wide array of states (for summary, see Keltner & Lerner, 2010).

There are active areas of research on at least 10 negative emotions: Anger, contempt, disgust, envy, fear, guilt, jealousy, sadness, and shame. And there are growing empirical traditions in studies of 11 positive emotions: contentment, enthusiasm, desire, love, compassion, gratitude, pride, awe, interest, amusement, and relief. As the empirical traditions focusing on these different emotions have matured, problems with early formulations of BET have arisen, as we detail here, in large part because theoretical claims that apply to 20 states (rather than 6 or 7 as was originally the case), necessarily encounter contradictions and qualifications. Generative theorizing has arisen that concerns itself with specific emotions (pride), or clusters of emotions (attachment related emotions such as love, desire, sympathy). This more local theorizing may represent the wave of the future of BET.

AS: A second clarificatory question concerns the role of context in your theory. A central theme emerging from Russell and Fridlund’s manifesto pieces is that context plays a key role in determining which emotions are perceived on the basis of faces, since the same face is interpreted in a variety of different ways depending on the context. What role does the context play in your multimodal proposal? For example, do you expect that context will trump the face+posture+voice combo the way it seems to trump the face alone, and if so what is the point of studying face+posture+voice combos in isolation from the context?

K-C: We believe that the ability to recognize an emotion from patterns of nonverbal behavior is only enhanced when multiple modalities of communication are involved. Select studies lend credence to this claim, but more evidence is warranted. When we see the full bodied, multi modal response of fear, or embarrassment, for example, recognition rises.

In our own empirical work, we also align with Fridlund’s BECV, in that we think that the signaling of intentions is critical to the occurrence of the emotion-specific expression. For example, in recent work we have shown that: a) displays of embarrassment signal the commitment to social norms to observers, so critical to the appeasement function of that display; b) displays of love signal a long term commitment to a romantic partner (Gonzaga et al., 2006). But we would suggest that signaling intentions and the underlying experience of an emotion tend to be correlated in the occurrences of specific emotions, that these two components of emotion are not independent, but most typically correlated.

What role does context play in our revised BET? Most importantly, we believe that the relational context between
signaler and perceiver—something first raised by Fridlund in his earlier work — leads to variation in the expression of emotion, and the inferences drawn by perceivers in judging the expression. For example, it is known that expressions of embarrassment signal more submissive social status. We would expect an embarrassment display to not signal such submissiveness when displayed by a high status individual. This sort of relational context variation in the meaning of emotional expressions needs to be systematically modeled, and is certain to reveal that the nature of the signal produced and the inferences drawn from that signal vary as a function of features of the relational context – power dynamics, degree of acquaintance, adversarial versus cooperative intentions, etc.

How much will the meaning of multi-modal patterns of expression vary as a function of context? The easy answer is that this is an empirical question. A safe bet is that context effects, which we grant are powerful, will be attenuated when multimodal signals are judged, given that they are likely to prove to be more reliable signals of emotion even in isolation from context.

AS: Let me move on to a couple of clarificatory questions on Fridlund’s manifesto piece. Alan, in your manifesto piece you reiterate the central proposal of the Behavioral Ecology View, namely that the face does not automatically express emotions but rather conveys social intentions, because automatically expressing emotions would contradict basic principles of animal communication. But why do we have to pick and choose? Since on a variety of theories of emotions, including BET, emotions involve states of action readiness, wouldn’t a face that expresses emotion *ipso facto* convey a social intention? For example, it is true, as you point out, that a smile can sincerely convey readiness to affiliate without the agent being happy, but if a face conveys that the agent is happy, why would it not also convey that the agent is ready to affiliate?

Alan Fridlund (AF): It’s worth noting, historically, that the term “emotion” was used to set off those behaviors that we couldn’t help, that “overcame us” or “got the best of us” – our passions (see Plamper, 2012, for the evolution of the term). “Emotion” was set against “motivation,” and motivated behavior was instrumental, feedback-guided, goal-oriented, strategic. “Emotion” was what was *un*intended.

Theories of emotion have metastasized, and the concept of “emotion” has now become so malleable that it can be injection-molded to inhabit any theory, even mine (see discussion by Fridlund, 1994). Thus, if emotions “involve states of action readiness,” then “emotion” can stretch even further to “include behavioral intentions.” Be my guest. Of course, all this wordplay hinges not only on what is meant by “emotion,” but also by “involve” (do you mean cause?), and by the ever-oxymoronic “state of action readiness” (is a state of readiness-to-act also a state of readiness not-to-act, and isn’t this true of all states, and a truism about readiness itself?). We have come so far and yet not moved at all.

BET formerly espoused the early “passion” view of emotion; the Facial Affect Program fired off patterns to the face keyed to our emotions, and those patterns might *once* have been indicants of our reflexive predispositions (that’s Darwin’s Principle of Serviceable Associated Habits, which Ekman, 1972, endorsed in his Neurocultural Model), but they are now vestigial, and we need cultural display rules to mitigate them.

This view changed when Ekman began embracing the Tooby-Cosmides “solving life tasks” emotion model in the 1990’s (cf., Ekman, 1992). At this point, emotions were no longer vestigial reflexes; they were organized, strategic motivations that arose in our Pleistocene “Environment of Evolutionary Adaptedness” (from J. Bowlby; see Tooby & Cosmides, 1990), and may or may not be suited to events in the contemporary world.
You ask if a face “conveys that the agent is happy, why would it not also convey that the agent is ready to affiliate”? I don’t accept that faces “convey happiness,” because I do not know a method of confirming scientifically the presence and the absence of happiness to allow this determination (and operationally, we need both). Nor do I believe that the face “conveys the readiness to affiliate” as a TV station conveys a storm warning or a car ad. Rather, a certain display, such as a smile, might probabilistically promote affiliation in a certain social context. By virtue of that evidence, the display may be labeled an “affiliative display,” not necessarily due to anything intrinsic to its neurology or its morphology, but due to its function as observed in social interaction.

AS: Alan, your theory of facial displays has been very successful in sensitizing the emotion community to the fact that facial displays communicate about social motives in an audience-dependent fashion, a view that, as Keltner and Cordaro’s manifesto paper shows, is now making its way also into BET. But I wanted to ask you for more clarification as to why it is evolutionarily advantageous to convey one’s social motives.

In a paper on animal communication that you often cite as a source of inspiration for your own views on facial displays, Dawkins and Krebs have stated the following: “It is information about what the actor will do next that poses theoretical problems…First, there is nothing to prevent animals ‘lying’ about what they will do next…and secondly, for an animal to declare its intention early on in a contest is equivalent to a card player showing his hand to an opponent at the start of the game. It is hard to imagine how selection could favour such behavior; instead, one would expect animals to conceal their eventual intentions until the last possible moment”.

My question is: if it is not to the evolutionary advantage of senders to broadcast information about their own emotions, why is it to their advantage to broadcast information about their own intentions? Also, is your skepticism about the viability of scientific definitions of emotions the primary reason why the Behavioral Ecology View does not allow emotions to figure in explanations of the origin and current function of facial displays?

AF: BECV does not hold that faces signal irrevocable intentions, just the possible next steps in on-going negotiation. Thus, a “threat face” would not signal that I’m going to attack no matter what – in that case, a sneak attack would be best – but that I may attack if you do not alter your behavior.

More generally, BECV argues that, depending upon the context, signaling one’s likely intentions can be an economical and safer alternative to enacting them immediately, with receivers countersignaling their own likely intentions based on their assessment of those signals. Both the signals and their assessments form the dynamic interchange by which interactants negotiate and determine their social trajectory (also see my manifesto).

As to emotion, making causal something that is ineffable is a fool’s errand. But that’s not the reason that “emotion” doesn’t figure in BECV; it’s that BECV’s functionalist view of displays neither needs nor benefits from “emotion” as an explanatory construct.

For BECV, the claim, “it’s not to the evolutionary advantage of senders to broadcast information about their own emotions,” is not only unverifiable but nonsensical (“information” is broadcast about emotions?). Something akin to that claim lies within BET, which maintains that display rules had to evolve culturally to mitigate inherited expressions, either because those expressions were vestigial reflexes (Darwin, 1872, and Ekman, pre-1992) or motivated acts best suited to an earlier era (Tooby & Cosmides, 1990; Ekman, 1992, and thereafter).

With regard to the ontology of emotion itself, I am sympathetic with Russell’s constructionist approach (he has
summarized it tidily in this debate), largely because of its humility. I see it as a springboard to anthropological investigations as to how diverse peoples may or may not “package” their behavior and experience into entities that may, or may not, resemble our Western emotions.

The cross-cultural studies on color and form (Gestalt) perception offer templates for understanding our tendencies toward packaging behavior into things like “emotions” as Kantian categories or Gibsonian affordances.

**AS:** Let me move on to a couple of clarificatory questions on Russell’s manifesto piece. Jim, in your piece you point out that happy people do not always smile, and that smiles occur without happiness. This is compatible with smiles having multiple causes: sometimes happiness, sometimes things other than happiness (you present several alternatives you think should have their day in court). But you also suggest that we “have no convincing evidence that emotions cause facial movements”.

This seems to imply that when people smile, we have no good reason to presuppose that happiness is *ever* (as opposed to always) the cause of their smiles. Is that your view, and if so what grounds your skepticism about the claim that happiness sometimes causes smiling? More generally, what sort of evidence would convince you that emotions sometimes cause facial movements?

**Jim Russell (JR):** At the heart of BET was a simple idea: feeling happy makes you smile, feeling fear makes you gasp, feeling disgusted makes you wrinkle your nose, and so on. Ekman (1980, p. 7) wrote, “When someone feels an emotion and is not trying to disguise it, his or her face appears the same no matter who that person is or where he or she comes from.” Ekman allowed that there are other causes of any facial expression (such as a display rule requiring a smile when greeting a stranger or when having your photograph taken). All the same, he implied that feeling happy, unless disguised, is sufficient for smiling. Ekman’s claim has not stood up well to the evidence.

The claim is difficult to test for various reasons, one of which is that emotion is typically confounded with other possible causes of the facial behavior. So, the scientific question is whether the emotion can be shown to cause the predicted facial expression when disentangled from other possible causes.

Consider the research program of Jose Miguel Fernandez Dols and his colleagues on happiness and smiling (e.g., Fernandez Dols & Ruiz-Belda, 1995; Ruiz-Belda, Fernández Dols, Carrera, & Barchard, 2003; Crivelli, Carrera, & Fernandez-Dols, 2015; Fernández-Dols, Carrera, & Crivelli, 2011; for general review, see Fernandez Dols & Crivelli, 2013). In a series of field studies, they found instances of intense happiness (such as winning in sports or orgasm) that could be disentangled from other plausible sources of smiling and in which attempts at disguise were unlikely. Intensely happy people rarely smiled, except during a social exchange.

So, we have no evidence for BET’s hypothesis, that is, no evidence that feeling happy, unless disguised, is sufficient for smiling, that feeling happy automatically causes smiling. You asked what sort of evidence would be convincing for BET’s claim. Had these studies yielded results in the opposite direction, the case that happiness per se causes smiling would have been supported. The available evidence indicates instead that a more fruitful line of research would view smiling as a social tool. Research could then pursue questions about the uses of that tool. Research on other emotion-face associations found similar results (Reisenzein, Studtmann, & Horstmann, 2013).

Your final hypothesis (“emotions sometimes cause facial movements”) is substantially weaker than Ekman’s statement (which did not include the word “sometimes”), fails to specify when those times would be, and is insufficient to support BET’s evolutionary story that facial expressions, as automatic unbidden signals, are parts of pre-prepared solutions to recurring situations from the Era of Evolutionary Adaptedness. A hypothesis about
“sometimes” can survive repeated failures to find supportive evidence, and “facial movement” moves away from BET’s specific predicted facial configurations as pre-interpreted signals.

A deeper problem with your hypothesis arises when “emotion” is defined, as it sometimes is, as a package that includes a facial expression as a part. Most affective scientists now agree that an episode of fear, for example, has multiple components. To illustrate, consider a case of fear that includes BET’s predicted facial expression: suppose that Sally undergoes an episode of fear one Saturday morning: she sees a bear, she appraises the bear as dangerous, her heart races, she looks for an escape route, she subjectively feels afraid, and her face at some point resembles the “fear face” (widened eyes and raised eyebrows) as BET has portrayed it. On the “package definition,” fear just is this whole package of components. What would we mean by saying that emotion, or fear, caused Sally’s facial expression? When we define “fear” as the full package of those components, one of which is the facial expression, we would mean that the whole causes its own parts.

On the “package” definition of fear, a more useful approach would be to ask if any of the non-facial components caused the facial component. This is the approach taken in a variety of competing theories. For example, in appraisal theory, specific appraisals produce specific facial movements (Ortony & Turner, 1990; Scherer, 1992; Scherer & Ellgring 2007). A similar approach is taken in psychological construction, which focuses on the components and seeks empirical evidence of relations among them.

Questions would be of this sort: Did her appraisal of danger cause Sally’s facial expression? Did her conscious experience, her feeling of fear, cause her facial expression? Did the massive increases in Sally’s sympathetic nervous system activity cause her facial expression? Did her visual scanning of the environment for an escape route cause her facial expression? All components will not be found in all (or even most) cases of fear, but the components recur, sometimes during fear but at other times as well. Each component can be studied separately for its relation to facial movement.

My guess is that the visual scanning of the environment is the most likely cause, and if so there is a ready functional explanation. On the other hand, one interpretation of Ekman’s (1980) statement quoted above is that the felt experience of fear causes the facial expression. Fernandez Dols and colleagues created a situation in the laboratory in which participants reported experiencing fear (while watching a clip from a horror film), but did not search for an escape route. These participants rarely showed the BET “fear face” (Fernandez Dols, Sanchez, Carrera, & Ruiz Belda, 1997). If evidence supports my alternative hypothesis about visual scanning, we would then have a precise account of the production of BET’s “fear face:” we would understand why it occurs sometimes during fear, sometimes fails to occur during fear, and sometimes occurs without fear.

An alternative to the “package” definition has been to think of “fear” as the cause of the components, with the emotion then presumably a specific pattern of brain activation. Of course, the brain causes the occurrence of each component, including any facial changes. But now the question is whether the same brain pattern causes all and only episodes commonly labeled “fear.” Is “fear” a useful scientific category for neuroscience?

For various reasons (heterogeneity, culture/language-specificity, historical shifts, fuzzy borders, and prototype structure), I argued that “emotion,” “fear,” and the like are not good scientific terms for psychology. For the same reasons, plus the fact that our linguistic ancestors knew next to nothing about what the brain does (Aristotle thought the brain’s function was to cool the body), I doubt that these terms are good scientific terms for neuroscience. I also note that some neuroscientists are abandoning thinking of emotion-specific brain circuitry (LeDoux, 2015).
AS: Let me briefly follow-up. Your answer suggests that you mostly oppose what we may call a deterministic notion of causation, according to which effects mandatorily follow from their causes. Would you be open to probabilistic notions of causation, which are widely used in other areas of science? For example, when an epidemiologist says that smoke causes cancer, she does not mean that cancer follows smoking mandatorily, but only that the probability of cancer is raised by smoking. Would you be willing to say that happiness is a probabilistic cause of smiling in the sense that it raises its probability? I suspect that many theorists who posit the existence of a causal relation between emotions and facial displays presuppose precisely this probabilistic notion of causation rather than the deterministic one.

JR: I am “willing to say” whatever the evidence requires. We can also speak of “probabilistic causation” if you like, although I know of no statement of BET in which the emotion–face link is labeled probabilistic causation. I view probabilistic causation as a way-station toward a scientific analysis, which would seek the mechanism involved. Early epidemiological studies found a correlation between smoking and certain cancers, but those studies did not answer the vital question of causality. Smoking might have been merely correlated with the actual cause, or early stages of cancer might have caused tobacco craving. Laboratory studies with non-human animals showed that something about smoking was indeed causal.

Two follow-up questions were then essential: What about smoking causes cancer? And, under what conditions does smoking cause cancer? Further research isolated specific carcinogenic chemicals in smoke and certain genetic predispositions. Because tobacco smoke is chemically complex, because cancer is a family of diseases, and because the mechanisms are not yet understood, the full story is far from complete today. The phrase “smoking causes cancer” is more a warning to the public than a scientific analysis.

The claim that emotion is a “probabilistic cause” of facial movement confronts the problems described above when emotion was defined either as a package or as a brain pattern. But, let us set aside worries over definitions and also suppose that happiness and smiling (or fear and the “fear face”) are correlated. The first question again would be whether something about emotion causes facial behavior. If evidence were affirmative, it would establish little other than it would be worthwhile to carry out closer, more rigorous experiments. The essential follow-up questions would then arise: what about happiness causes smiling, or fear the “fear face”? Under what conditions does happiness cause smiling or fear the fear face? That is, which components of the emotion are causal and in what contexts?

I think that my answer can be translated into your terms. If, as BET hypotheses, under conditions with no motive to disguise, happiness per se automatically results in smiling, or fear the “fear face,” then the link would be deterministic causation. If happy people often smile, but sometimes not, and if scared people often show a “fear face,” but sometimes not, then the link would be correlational. If further evidence established that something about emotion was causal, you could call it probabilistic causation, but we would need to search for what about the emotion is causal when it is and under what conditions. As in the fictional example above of Sally’s fear, some answers would be compatible with psychological construction, some would not. I’m not sure what BET predicts exactly. Still, to repeat, I know of no evidence that when confounds are eliminated, emotion (however defined) is causal in any sense. If such evidence were to appear, then to me the follow-up questions would be of great interest.

AS: Jim, in your piece you emphasize the degree of variability characteristic of the facial expressions correlated with any discrete emotion. Basic emotion theorists try to explain variability by invoking display rules, blending between different emotions, the fact that a certain intensity threshold may have to be cleared before the relevant expression is manifested and the fact that basic emotions form families. Do you
consider these sources of variability at all legitimate, or are they in your view simply ways to preserve the theory from falsification, as you suggest at least in the case of display rules?

JR: Perhaps the most surprising empirical finding in affective science is the amount of variability found within any particular emotion category. There is huge heterogeneity within the category, say, of fear or anger. The components of the emotion are less tightly joined, less correlated with each other, than anticipated. One instance of this general finding is what you term “variability characteristic of the facial expressions correlated with any discrete emotion” – by which you mean, I assume, that the predicted facial component is less correlated than anticipated with the various non-facial components of a given emotion.

Because BET assumed that, for a given emotion, the components are highly intercorrelated (since they all have the same cause, activation of the “affect program”), the theory faces a challenge in explaining the low correlations. You listed some attempted explanations. Display rules, blends, thresholds, and a family structure on their own are all sound hypotheses, but problems arise when these hypotheses are joined with BET.

Klineberg (1938) pointed out that societies have rules about faces (in my society: downcast eyes at funerals, smiles at parties), and I believe that people often follow these display rules. Ekman (1972) invoked those rules to explain how, given the same emotion, different facial expressions can be observed in different societies. However, when an emotion occurs without BET’s predicted facial expression and a display rule is invoked as the explanation ad hoc, then BET’s analysis becomes immune to evidence.

Consider, for example, the highly cited Japanese-American study on display rules. American and Japanese participants showed similar facial movements to a disgusting film during a private viewing, whereas they showed different facial movements to the film in a social situation. Ekman (1984) summarized: “In private, when no display rules to mask expression were operative, we saw the biologically based, evolved, universal facial expressions of emotion. In a social situation, we had shown how rules for the management of expression led to culturally different facial expressions.” No particular display rule had been specified ahead of time, and thus no prediction as to what facial behavior the rule required. No evidence was offered that Japanese are subject to a display rule and Americans not. No evidence for the operation of a display rule was offered other than the lack of BET’s predicted facial expression. No evidence was offered for the non-operation of a display rule when the predicted facial expression occurred during the private viewing, other than the occurrence of the expression.

A display rule was offered as the explanation of the observed cultural difference in the social situation, but alternative explanations, such as differences in emotion or in focus of attention, were not ruled out. (Incidentally, I followed Ekman in assuming that BET’s predicted facial expressions occurred during the private viewing, but it is not clear that this assumption is correct.) See Fridlund (1994) for a revealing analysis of this study and the display rule concept more generally.

The other hypotheses that you mentioned (blends, thresholds, and families) all use the concept of “emotion” to explain failures of BET’s prediction. As I said, the heterogeneity, culture/language-specificity, historical shifts, fuzzy boundaries, and prototype structure of the concept of “emotion” make it a poor tool for science. Adding blends, thresholds, and families complicates matters even further and potentially adds to BET’s immunity from evidence.

I know of no version of BET that includes a worked out account of blending or thresholds. If fear is blended with anger, what happens? Do eyebrows go up or down or cancel? Is the behavior attack or flight? Or attack while fleeing? What happens in the autonomic nervous system? (Levenson, Ekman, Heider, & Friesen, 1992, found
finger temperature higher in fear but lower in anger. Again cancel?) What was the adaptive value of such a blend during the Era of Evolutionary Adaptedness?

The lack of predictive rules for blends increases the danger of making the hypothesis ad hoc (blending produces whatever is observed). Similarly, I know of no explicit hypothesis stating threshold values nor any means to measure the intensity of the entire emotional episode. Intensity of each component is potentially meaningful and measurable, but I know of no means of integrating intensity across components or whether the concept is meaningful. Are all components weighted equally? What is the intensity of an emotional episode with a frown, extreme sympathetic activation, but no vocal change and a denial of any felt emotion? Does it pass the threshold? What about the reverse: no frown, no sympathetic activation, but vocal changes and a claim of extreme emotion? (If BET considers these two cases as the same category of emotion at the same intensity, it would predict similar consequences. Psychological construction, in contrast, considers the components different and therefore anticipates different consequences.)

That a category of emotion consists of a family of events is similarly a sound idea. Long ago, Beverly Fehr and I (1984) offered evidence that the concepts of emotion, anger, love, and so on are better understood from a prototype theory of categories than from the classical theory of necessary and sufficient features. I mean that the category has fuzzy boundaries and an internal structure in which some cases are better exemplars than others. Rather than sharing common features, exemplars within the category share a family resemblance. One implication of these findings is that category membership is less predictive than it would be if the category were classically structured. That category members form a family rather than a precisely defined set is a property of the category in the mind of the categorizer, not of the external world. After all, the categories of witches, superheroes, and goblins similarly show prototype structure. Family resemblance rather than common features makes emotion categories poor scientific tools and is therefore a challenge to BET.

Finally, let me return to the basic finding that initiated this question: the components within an emotion category are less intercorrelated than anticipated. Cases share a family resemblance rather than common features. This finding is a challenge for BET, but not for other conceptual frameworks. In psychological construction, for example, emotion is replaced with its components, with no assumption that they stem from the same cause or are highly intercorrelated. Most of the components are neither emotional per se nor produced by emotion-specific processes. Certain components under certain circumstances might well cause facial movement. For instance, if a component of some instances of fear is visual scanning of the environment to locate an escape route, then visual scanning in those instances can cause eyes to widen and brows to rise. But then visual scanning, even without fear, would have the same effect. Fear, without visual scanning, would not.

AS: In this final stretch of the debate, I would like to ask three further questions to all debaters. The three manifesto pieces seem to suggest that all parties at this point agree that there is no one-to-one correspondence between facial displays and discrete emotions such as anger, fear, disgust, happiness, sadness and surprise, and that the snapshot method that relies on prototypical facial displays is fundamentally flawed. Is this indeed your view? If so, can this new area of agreement generate further convergence between research programs on facial displays, and if so how? Finally, what do you suggest as the best alternatives to the standard snapshot method using prototypical facial displays for the study of facial displays?

JR: Indeed, it is my view: There is no one-to-one correspondence between facial displays and discrete emotions, either for production or interpretation. And the snapshot method that relies on prototypical displays is fundamentally
flawed for the purpose for which it has been used.

I anticipate further convergence. I applaud Keltner and Cordaro’s emphasis on voice and body as they change over time as further sources of information that an onlooker uses and on their expanded list of inferences the onlooker might make. Their multimodal expansion converges with traditional accounts in the study of nonverbal communication, including my own. I hope we can continue the expansion on both the expresser’s side and the onlooker’s side.

I suggest that the expansion continue by including the expresser’s physiognomy, current situation, the broader context in which the facial change takes place, and the interaction between expresser and onlooker. Indeed, there is likely no information about the expresser that an onlooker might not use. We can also expand the list of inferences the onlooker makes; the onlooker tries to infer the expresser’s character, cognitive state, and likely behavior. Indeed, the onlooker can infer an uncountable number of things beyond characteristics of the expresser. (If I see the “expresser” squinting while facing West, I might guess the time of day.)

Because of the complexity and dynamic quality of this process, the information relied on by the onlooker is not well captured by a snapshot of the expresser at one slice in time. The process is an on-going rapidly changing interaction with a history and a future. In acknowledging this dynamic quality, Keltner and Cordaro move toward Fridlund’s account, at least a little. In my lab, we have recently taken a step in line with Keltner and Cordaro’s expansion by including face, voice, and body posture, all changing over time (Nelson & Russell, 2011 a,b,c; 2012; 2013). We need to think of interpretation as a process rather than a labeling. New methods will have to contend with much more complexity and with the interactions involved.

As the processes of production and interpretation are empirically examined, the results will force theoretical convergence. I’m optimistic that researchers will converge on an account of facial behavior and its interpretation – just as convergence can be seen on other aspects of emotion. For example, consider the notion that the emotion as felt plays a causal role in certain facial movements. To examine this idea, we need an account of the felt emotion, and there is convergence recently. In my account, the emotion as felt (the subjective experience of feeling a specific emotion or what I called emotional meta-experience) is cognitively constructed.

I see no difference between my constructionist account and that offered by Levenson (2011), a prominent BET theorist. And the same basic account of the felt emotion was voiced by LeDoux (2015) from a neuroscience perspective: “I thus agree with basic emotions theorists that some responses that occur when people feel emotions are hard-wired, but I don’t think that the emotion, the feeling, is hard wired. It is imputed, assembled, constructed, or otherwise cognitively created.”

**K-C:** We think it is a stretch to say that the snapshot method is fundamentally flawed; rather, it was a first attempt in a young science to understand how emotions are expressed, and it has generated enormously useful evidence and stimulated provocative debate, as evident in this exchange. The snapshot evidence – the matching of emotion terms to static, prototypical expressions – has produced two fairly entrenched camps in the field, one that presupposes that those static expressions do say something significant about emotional expression (us), and another that suggests that there is little meaning in those displays (Russell, Fridlund). Given this state of affairs, perhaps a way out of the stalemate is to turn to new empirical approaches. That was our hope in writing the manifesto piece.

Our review was founded on converging empirical discoveries over the past 15 years showing that the snapshot
method only captures a small portion of the meaningful variance in a) how people express any particular emotion, and b) how people recognize emotion in other people. This is far from surprising: snapshots of facial muscle movements ignore other modalities of expression, and they represent a split millisecond in time, ignoring the dynamic properties of emotional expressions, now of increasing focus in the field. The critiques of Russell and colleagues point to other important sources of variation in need of theoretical modeling – contextual information, variation in the meaning of emotion concepts across individuals and cultures, and so on. So we agree that we should move beyond the classic snapshot, emotion recognition approach, but note that such approach, flawed as it may be, was pivotal in getting the science of emotion expressions off the ground.

With respect to new approaches and possible convergence, our version of BET has deep ethological foundations – the study of naturalistic behavior in real world contexts – and it bridges various gaps with alternative experimental approaches. First, and this comes straight out of our review and the ethological foundations of BET, we recommend looking at multi modal, dynamic expressions of emotion. A new generation of recognition studies is needed, which considers the signal value of multimodal, dynamic patterns of behavior, while taking to heart the critiques of forced choice methods and the need for free response data. We think this will begin to capture individual and cultural nuances in interpretations of emotional expression.

Second, we should study emotion in the contexts in which it is most likely to occur, which, as Fridlund has noted wisely, is in social contexts. Most of the canonical emotions that attract most of the scientific attention are social – anger, embarrassment, pride, love, sympathy, amusement, etc. We think that research will be well served by studying people in relationships in actual social interactions, rather than in sterile laboratory circumstances responding to static images and film clips. And in this vein, with all of the real world data now available through Youtube and other new social media, there are intriguing opportunities to study expressive behavior.

A quick search of videos can provide real-world evidence of what happens when people are in truly terrifying circumstances, or in real adversarial encounters with enemies, or engaged in genuine celebration. So we would suggest that the field move more closely to the more ecologically valid elicitors of emotion as a means to capture expressions of emotion. Such methods could be used to capture context-specific expressions, and what observers infer from such expressions. The recent work of Tracy and Matsumoto on shame and pride displays following losing and winning in the Olympics is illustrative of such an approach (Tracy & Matsumoto, 2008).

AF: This question stipulates the existence of “discrete emotions,” implies the existence of “prototypical facial displays,” and asks whether I assent to the proposition that the two are not linked. To me, this is like asking whether I indeed agree to no relation between phlogiston and dragons.

Let me open up the question so that it is not an epistemological trap: I do not believe that there are consensual scientific definitions of “emotion” that would allow us to know when a person is having and not having an emotion (and by Keltner and Cordaro’s account, “feeling” is now optional, which knocks self-report data off the table). Furthermore, we have no such definitions for the specific emotions referred to by the terms “happiness,” “sadness,” etc. Second, I do not believe that there are “prototypical facial displays,” and because we have no workable definitions of emotion, we cannot produce meaningful data about the relations between the two. To say that the “snapshot” method is flawed would be to suggest that the wider presumptions aren’t equally flawed.

AS: Let me briefly follow-up on your answer for further clarification, Alan. First, when you say that my question “stipulates the existence of discrete emotions”, are you suggesting that you are not convinced that discrete emotions exist, in the sense that anger, fear, disgust and so on are fictions like dragons and
phlogiston? Or are you simply questioning that anger, fear, disgust and so on are suitable categories for scientific investigation, on account of the difficulty of providing a consensual definition for them? Second, once we grant that the snapshot method is flawed both in its specifics and in its wider presumptions, the question of how to move forward remains open. So I will ask you again: What alternatives to the snapshot method do you consider most promising for the study of facial displays?

AF: As a philosopher, you know that asking me to declare whether “discrete emotions” such as “anger, fear, and disgust” actually “exist” is absolutely meaningless unless you define the terms before you ask. Specifically, one must say what is meant by “discrete emotions,” and then what it would mean for them to “exist.”

Let’s keep it simple. Discrete emotions such as anger, fear and disgust certainly exist as concepts for BET advocates, and probably for a good number of ISRE members. To be formally philosophical, I agree entirely with that claim and the sentence containing it. Can anything more be demonstrated scientifically? Perhaps on Twin Earth? That’s where so much of this debate has become mired.

For BECV, the essential question, thankfully, is not about whether “emotions exist,” but “How do our facial movements affect the trajectories of our social interactions?”. This question precludes snapshots, because our social interactions are diachronic, cumulative and interwoven. Specifying contexts (our interaction contexts are cumulative, and begin to comprise relationships), and capturing and understanding the different ways that people move their faces and how their interaction partners react, requires a dynamic, functional view, not one constrained a priori by BET’s freeze-framed boxes. Chovil’s (1991) inductive typology of communicative facial displays is one way to advance.

AS: My second question concerns the issue of universality. All parties seem to agree that the question of whether a certain emotion is or is not recognizable – an either-or distinction – from a facial display is not useful and that there are cross-cultural differences in the degree to which facial displays are recognized. But then debaters part ways on whether anger, fear, embarrassment, disgust, embarrassment, pride and so on are in any interesting sense universal.

Keltner and Cordaro seem to suggest that we will find universality at the level of multimodal and extended expressions rather than unimodal and momentary expressions, with the proviso that there will be gradients of recognition. Russell suggests that the better-than-random ability of recipients to match pictures with emotion labels does not establish any interesting form of universality. Fridlund doubts that recognition studies can help us settle the question of the roles of biology and culture in the production of facial displays, because species-constant learning can explain evidence of cross-cultural recognizability just as well as natural selection. Could you state your final position on universality, describe its implications for the study of facial displays, and briefly summarize what you make of the views on universality expressed in the two manifesto papers other than yours?

K-C: Our final position on universality is that, as the methods and measures become more refined, and researchers move to more multimodal assessments of emotion, researchers will continue to document a substantial degree of universality. This work is inordinately hard to do, given how difficult it is to identify emotion-eliciting stimuli with similar meanings across cultures, the cultural variations in the meaning of words, and the reluctance of the field to study actual behavior. This being said, we expect different emotions, as postulated by BET, to have some core signal behavior, broad patterns of emotion-specific physiology and underlying appraisal, and similarities in the functional outcomes of different expressions (e.g., anger expressions deter observers of undesirable behavior).
Ultimately, we agree with Fridlund, that it will be hard for expression-related data to fully tease apart the biological and species-constant learning explanations of universals in emotion recognition, all the more problematized by today’s technologies of smart phones, Facebook, and emoji. We believe the field will have to turn to other measures – genetic markers, peripheral physiology, immune system response – to generate more unequivocal support for the thesis that emotions are evolved, genetically encoded biological processes that are observed in different cultures – the core of the universality thesis. We hope that our review points to several intriguing possibilities – the blush response, vagus nerve activation, the cytokine response – that await empirical study and that are open to falsifiable hypothesis testing.

**AF:** “Recognizability” stipulates what is only a BET presumption: that faces express universal emotions, with the only issue being the extent to which observers can divine them when they see them. The alternative, of course, is that faces move in all kinds of ways, and observers ascribe qualities to, and make inferences about, the faces they observe. If my wife comes in from outside and I ask her about the weather and she scowls, I see from her face that it’s lousy out. At that moment, in that context, her face expresses what the weather is like.

The varieties of information that others can ascribe to the face are nearly limitless. MIT mathematician Herman Chernoff used the individual features of cartoon faces as variables to represent arbitrary sets of multivariate data such as demographics and, yes, weather (Chernoff, 1973; Fridlund, 1994).

Chernoff was onto something. Here’s a photo of a Papua New Guinea woman I pulled off the Web:

What is she “expressing” with her face? Let’s imagine the possibilities. Suppose I just asked her:

- What’s the food like in the tavern down the road?
- How constipated are you?
- What do you think of the village leader?
- Is your husband faithful?
- Do you like me taking your picture?
- Was your son sick for very long before he died?
- How’s the medical care in your village?
- Would you want to travel to Mars?
- Do you think it was the monkey meat or the melon?
- The woman down the road just told me she was gonna kick your ass.

So her face can “carry” information about her colonic health, the status of her village’s medical care, and possibly her views on interplanetary travel. The face can “carry” information about nearly anything, really. In each case, “carry” only means move within a context, with observers making ascriptions to, and inferences about, the movements they observe within that context.
If facial "recognition" is presumptive, then Keltner and Cordaro’s "recognition gradient" is even more so, because they observed neither recognition nor a gradient. They merely obtained a random assortment of matching-to-sample scores, and arranged them from maximum to minimum. Mathematically, a gradient is a continuous function of Y on X, in the form of \( Y = f(X) \), and it requires two scalar variables. There is an inverse-square gradient of light by distance, for example.

Here is an analogy to what Keltner and Cordaro have shown. It’s a graph of annual temperatures (Celsius) for 12 of the world’s major cities, as plucked from a table on [www.statcan.gc.ca](http://www.statcan.gc.ca). Before I graphed them in a bar chart, I sorted the cities from highest to lowest average annual temperatures. Can I claim that I have discovered a “temperature gradient” for major cities of the world? No, I can only claim to have found an assortment of cities that differ in temperatures, and that I know how to sort the temperatures from highest to lowest.

Getting a gradient requires two scalar variables. In this case, the abscissa needs to be scalar and not categorical; this mandates a regression plot instead of a bar graph. How might I proceed? I might have a hunch that distance from the Equator has something to do with the annual temperatures in the original table. Therefore, I found the equatorial distances in km (using [dateandtime.info](http://dateandtime.info)) for the 12 major cities and plotted them against the annual temperatures:
Now we have an actual gradient, of annual temperature as an (apparent) inverse function of equatorial distance, with each of the major cities marked on the regression plot. There are two outliers, which requires further explanation (air pollution?).

Had Keltner and Cordaro framed and tested an hypothesis, e.g., that the matching of certain facial or face + body configurations might be a function of Westernization or the use of an Indo-European language or some other parametric variable, they would have been able to claim a “gradient.”

Gradients of matching-to-sample scores for multimodal expressions, as Keltner and Cordaro propose, will require multivariate regression plots in $n$-space, where $n$ is the number of variables that will constitute their expressions, and each variable of $n$ is a time-series vector given that Keltner and Cordaro wish to measure dynamic expressions. Culture samples will need to be compared for similarities and differences using confirmatory factor analyses and discriminant analyses. These techniques are difficult enough for “snapshot” data; they are nearly intractable for the time-series data which must be obtained. I know these issues all too well since, long ago, when I was a BET acolyte, I was among the first to use these multivariate techniques for the “recognition of emotion” (Fridlund & Izard, 1983; Fridlund, Schwartz, & Fowler, 1984).
Onto universality. What question do we want answered by pursuing the “universality issue”? If it’s the biology vs. culture conundrum, such studies will always be indeterminate, for the simple reason that both uniformity and diversity are expected products of either cultural or natural selection. Uniformity can arise phylogenetically or via convergent evolution (convergent cultural evolution was misnamed “species-constant learning” by Gordon Allport, and the term persisted with Ekman), making nonhuman primate studies equally indeterminate. It is worth asking, however, whether certain features of our facial displays vary with certain genotypes, as a function of known human migratory patterns (e.g., they may correlate with hemoglobin group dispersion; e.g., Fridlund, 1994), or are dependent upon the phonemic makeup or prosodic patterns of the speech they accompany across and within cultures.

I do not believe that Keltner and Cordaro have resolved the definitional issue with “emotion” but instead have complicated it. In their formative writings, and in the Neurocultural Model (Ekman, 1965, Ekman & Friesen, 1967, Ekman, 1972), Ekman and Friesen followed Tomkins in arguing that emotion was expressed strictly by the face, “while the body shows the adaptive efforts of the organism to cope with the affect state” (Ekman & Friesen, 1967, p. 718). Keltner and Cordaro now reverse that position by re-adding what Ekman and Friesen jettisoned, the very gestures and movements that foundational nonverbal communications researchers like Mead and Birdwhistell always considered important and modern ones still do (e.g., Bavelas & Chovil, 2000).

There are two problems with Keltner and Cordaro’s tack. The first is prosaic: adding variables – especially in the form of multivariate time series – complicates prediction and analysis, and leads to accumulation of error terms and the need for ever-larger sample sizes to achieve statistical power. The second is substantive. Keltner and Cordaro want to add gestural and other bodily movements to their “expressions of emotion.” By what procedures will they determine which movements, committed by whom and when, are part of the emotional expressions themselves, and which movements, committed by whom and when, reflect cultural display rules to manage those emotions or even fake them? In the original Japanese-American cultural display rules study (reported in part by, e.g., Ekman, 1972; and see Fridlund, 1994, for detailed description, including its never-published “missing” condition), self-report data were used to verify the presence of “emotion”, but what would be criterial for emotion now without qualia as definitive?

Keltner and Cordaro, by making qualia optional in their “New BET” definition of emotion, would be counterfeiting the term while profiting from its cash value among non-BET researchers who haven’t yet been let in on the secret. New BET researchers could claim to be studying “emotion” while they are really studying “emotion*”, which is emotion-without-the-feeling. Thus, in a New BET study that concluded, “happier people smile more,” the study would have to be footnoted, “NOTE: In accordance with the provisions of New BET, people who are happy do not have to ‘feel’ happy (Keltner & Cordaro, 20xx).” This will delight readers, I am sure – although they need not feel it.

I find Russell’s Minimal Universality position, which he reviews in this debate, quite compatible with my own thinking, because it presumes little and builds from the bottom. We know too little to continue to force-fit huge, lumbering theories onto scant data.

JR: In the phrase “degree to which facial displays are recognized,” you seem to assume that faces display BET’s specified emotions, but that should be a hypothesis rather than an assumption. Also, “recognition” is a loaded term because it presupposes that the emotion is there to be recognized or not. That too should be a hypothesis rather than an assumption. That “better-than-random” performance fails to establish BET’s claim is a simple matter of appropriate statistical inference.
I urge continued studies of universality, although not continued use of the same old methods. Studies of universality are also studies of differences among societies, and we cannot understand human psychology without observations on similarities and differences among societies.

On the production of facial movements, we have insufficient evidence to go on, but I guess that faces universally move primarily (but not exclusively) as part of three processes: (a) paralanguage (i.e., facial movements have semantic, syntactic, and pragmatic functions during speech), (b) perceptual-cognitive processes (e.g., widening the eyes in visual search, moving the lips in tasting, and so on), and (c) social negotiation (i.e., the social exchanges discussed by Fridlund).

On the interpretation of facial movements, everyone agrees that observers derive information from another’s facial movements. I proposed a form of universality I called “minimal.” People everywhere use the expresser’s facial movements, in conjunction with whatever other information is available, to infer what they can about the expresser (Russell, 1995). So, for example, categorization of the state of the expresser might be a universal process, but the system of categories used might vary somewhat with culture and language. People in modern Western cultures interpret one another by relying on, among other things, their system of categories that include “emotion,” “anger,” and so on, but people in cultures remote from the West do not have exactly the same categories.

I used the word “minimal” to characterize the small number of non-controversial assumptions needed to account for current evidence and to contrast with the maximal universality posited in BET: Ekman (1972, p. 279) wrote, “The evidence now proves the existence of universal facial expressions... Regardless of the language, of whether the culture is Western or Eastern, industrialized or preliterate, these facial expressions are labeled with the same emotion terms: happiness, sadness, anger, fear, disgust and surprise.”

On the other manifestos, I find Fridlund’s manifesto compelling, in its analysis of the nature-nurture issue, in its critique of emotion-based theories, and in its proposals for a new theory of facial expressions. I applaud Keltner and Cordaro for making their version of BET responsive to previous critiques. Their research will yield valuable information. Implicitly, they acknowledge a need to revise BET. I believe that a major challenge in doing so will be making the theory testable. I remain skeptical that BET can be revised to accommodate the evidence without abandoning its core tenets. Indeed, after this exchange, I wonder just what core tenets of BET remain. I hope to see empirical and theoretical pursuit of all three – and indeed alternative — conceptual frameworks.

**AS: My final question concerns what is arguably the most intractable difference between debaters, namely the question of whether or not the field of affective science has managed to provide a productive scientific definition of emotion and its subordinate categories. Dacher and Daniel seem happy with a modified version of Ekman’s (1999) original definition, according to which basic emotions are solutions to fundamental life tasks with 11 characteristics, none of which is a *sine qua non* for a basic emotion to be instantiated (with the possible exception of distinctive universal signals, whose *sine qua non* status is presented by Ekman as a “challenge”).**

The list includes distinctive universal signals, distinctive physiology, automatic appraisals tuned to distinctive universals in antecedent events, distinctive developmental appearance, presence in other primates, quick onset, brief duration, unbidden occurrence, distinctive thoughts, memories and images, and distinctive subjective experiences. Dacher and Daniel, could you explain why you think that this definition is promising, and if so how do you propose to improve on it in light of what has transpired from this debate? Jim and Alan, could you explain why you find this definition to be hopeless, and why you
believe that no amount of tweaking can make it viable?

**AF:** In recent years, BET has presented a moving target. One major change was prompted by the Ortony and Turner (1990) attack on “basic emotions.” Ekman responded by introducing the term “emotion families” (Ekman, 1992), which widened BET’s criteria for expression of any given emotion.

Ekman altered his writing about emotion itself by around 1992 (Ekman, 1992), when he began citing Tooby and Cosmides’s (1990) “life task” view of emotion as adaptation. This motivated and strategic reformulation of basic emotion theory is fundamentally incompatible with the Neurocultural Model’s (Ekman, 1972) earlier view of emotional expressions as Darwinian, vestigial habits that had to be managed by cultural display rules. The ramifications of this switch for the theory’s tenets are unacknowledged among BET advocates, because whereas the earlier version mandated display rules to govern expressions that were stipulated to be vestigial, the newer “adaptationist” version allowed that the expressions might have evolved on their own as components of behavioral patterns that enhanced inclusive fitness, with no requirement for cultural display rules to squelch or counter them on the fly.

Yet another radical change was in response to findings of implicit sociality in solitary facial expressions (e.g., Fridlund, 1991). Ekman and Keltner (1997) dismissed the findings summarily. Ekman then revised BET to encompass the notion of solitary display rules (Ekman, 1997) to explain solitary expressions, although this concept vitiated the rationale and conclusions of BET’s only empirical claim to display rules, the original Japanese-American study, which rested on the proposition that solitary expressions had to be spontaneous and “emotional” (Ekman, 1972; and my manifesto).

Finally, in a vast leap from his original view that emotion was categorical and decodable only from the face, Ekman (1999) provided a list of 11 characteristics, of which (quoting the above) “none …is a *sine qua non* for a basic emotion to be instantiated." What are we to do with this? A laundry list of vague qualifiers such as “unbidden occurrence” (to whom?), “distinctive thoughts,” and “brief duration” do not make for a scientific definition. What is of equal importance, but has been missing in every BET variant thus far, is the complementary set of exclusion criteria – a stipulation of those conditions under which emotion, or a particular emotion – is not occurring.

As I noted earlier, Keltner and Cordaro are now loosening the criteria even further by reclaiming within “emotional expressions” the bodily movements that the original Neurocultural Theory jettisoned, and it is reasonable to ask whether this sprawling theory, with its evasions, revisions, slipclauses, and patches, has any remaining coherence or explanatory value.

Historian of science Jan Plamper (2012) has surveyed the history of the BET enterprise, as critiqued in depth by Leys (2007, 2010), and declared that, “[f]rom the standpoint of the natural sciences and according to the principles of scientific procedure and verity, Ekman’s research is bankrupt” (p. 158). Plamper’s summary judgment applies to Ekman’s neurocultural version. He could not anticipate Keltner and Cordaro’s “new BET” effort to salvage BET by offloading some of its core tenets, which strikes me as trying to keep the BET ship aright by hacking away at its keel.

There is a larger perspective from which I disagree with BET. For me, BET is riddled with a preconception of emotions as hermetic “things” with inner circuits and outer manifestations (Leys, 2010), and these emotions in BET are intertwined with a Western, Augustinian conception of a “real self” versus a “managed self” (Cary, 2000; Fridlund & Duchaine, 1996). Combined with the fact that the questions and methods used to study expression
within BET are so Procrustean (readers will note the relentless BET framing of the questions even in this debate), and the predictions possible within BET are now so imprecise and slippery, little of substance will likely be learned.

For my money, we need more anthropologists and historians to acquaint us with the various ways in which the peoples of the world make – and made — sense of their behavior in their terms. As for understanding how we signal each other and navigate our social terrain, BECV offers a clear-cut, alternative direction.

**JR:** The word “emotion” points to real and important events and dispositions, which we need to study and understand. Still, it does not follow that “emotion” defines a scientifically tractable domain. Similarly, the word “constellation” points to real stars in a real geometric configuration from the perspective of earth. “Constellation” was once a term used in astronomy, but was later abandoned as a scientific term because constellations do not explain events of interest to astronomers.

Andrea’s final question raises the question whether “emotion” as defined by BET theorists is a scientifically tractable domain. Let me first ask, however, if Ekman’s list of 11 characteristics really is a definition and, if so, what sort of definition it is. Let me distinguish two types of definition, each serving a fundamentally different purpose (Russell, 1991; Widen & Russell, 2011).

What I call a “descriptive definition” aims to capture a folk concept: What do ordinary people mean by the word “emotion”? This is a topic in the semantics of modern English: how to articulate the meaning of a non-scientific concept in everyday use. Much as an anthropologist might try to capture what folks in a certain society mean by words translated as “demon” or “chief” or “spirit,” psychologists can try to define “emotion,” “anger,” and so on in a way that captures their everyday meaning.

Descriptive definitions of emotion terms are important because the concepts expressed play a role in the way folks think, remember, and judge propriety and morality. The everyday concepts of emotion have a role in law and are involved in communication between psychologists and the people being studied.

I doubt that Ekman meant to offer his list of 11 characteristics as a descriptive definition. I say this because, as an attempt to formulate a descriptive definition, the list strikes me as implausible. Each characteristic listed would have to have a psychological reality: ordinary people would have to understand and use each characteristic to decide what is and what is not an emotion. But, the characteristics are themselves in need of definition, and ordinary folk do not have access to some of them: distinctive physiology, automaticity of appraisals, universality of signals, presence in primates, and so on.

Indeed, the characteristics do not help us decide whether a specific event is or is not an emotion. If I have a certain subjective feeling and wonder whether it’s love or indigestion, how does “presence in primates” help me decide? Furthermore, to be a successful descriptive definition, the set of things to which the definition applies would have to overlap perfectly with the set of events named emotion in everyday thinking. But it is pretty clear that this is not the case. Consider brief duration. The set of events of brief duration leaves out love and hate, most cases of grief, anxiety, and depression, and some cases of happiness, sadness, fear, and anger. Note that this blocks Ekman’s strategy of handing out some items that do not meet the definition of basic emotion to a theory of emotional plots or other affective constructs. All “blue ribbon” basic emotion types like anger and fear have tokens that do not satisfy the definition.

One possibility is to consider some of the characteristics on the list of 11 optional. The problem is if we make too many characteristics optional we end up capturing items that do not have anything to do with emotions, from
angina to xenophobia. Angina lacks some of the characteristics but it includes others: distinctive physiology, quick onset, brief duration, unbidden occurrence, distinctive thoughts, memories and images, and distinctive subjective experiences. Xenophobia similarly lacks some characteristics but it includes others: automatic appraisals, quick onset, unbidden occurrence, distinctive thoughts, memories and images, and distinctive subjective experiences. Indeed, if all characteristics can be optional, as Ekman seems to suggest, it is hard to imagine a psychological event that is not included.

In any case, if interpreted as a descriptive definition, Ekman’s list would compete with other accounts (Fehr & Russell, 1984; Johnson-Laird & Oatley, 1989; Ortony, Clore, & Collins, 1988; Wierzbicka, 1999). Which of these accounts best captures the folk concept of emotion in everyday use is an empirical issue, and we would have to see evidence about how regular language users use emotion concepts to settle the matter.

What I call a “prescriptive definition” aims to provide scientists with a useful concept for understanding some scientific phenomenon. Much as physicists have modified the concept of force from its primitive folk meaning to its current prescriptive definition in physics, so psychologists can begin with the folk concept of emotion and modify it until it becomes a scientific tool. Scientists can stipulate what they mean by a scientific term as they build a theory of the phenomena. If “emotion” is to become a scientific tool, we would need to develop a prescriptive definition for it.

I doubt that Ekman’s list is meant as a definition in this prescriptive sense either. Prescriptive definitions used in science – e.g. the definition of probability as relative frequency, the definition of intelligence as ability to pass a Turing test – aim to increase precision and offer clear criteria of inclusion and exclusion. If Ekman’s list really were a prescriptive definition, it would have to dictate more precisely which cases are and which are not basic emotions. With no “sine qua non” (necessary) characteristic and no exclusion criteria, the set of events to be tested and explained is massively indeterminate and heterogeneous. As a result, the basic goal in science of making concepts measurable is off to a poor start.

Incidentally, I have never seen a research project based on this definition – that is, a project in which a sample of human events was objectively assessed on the 11 characteristics and those that pass some threshold are then stipulated to be emotional events, with the rest regarded as non-emotional events. Nor has this list, to the best of my knowledge, served as a prescriptive definition in the development of theory.

Let me suggest instead that Ekman’s list of 11 characteristics is a set of empirical claims. Research projects in BET implicitly rely on the folk concept of emotion (either through the experimenters’ choices or ratings by participants or observers) and view the list of 11 characteristics as a set of empirically testable hypotheses about the events picked out by the folk concept “emotion”.

Indeed, Ekman clearly treated each characteristic as an empirical finding. For example, the (alleged) fact that facial signals are universal was presented as a major empirical discovery. The discovery was that what we call anger in ordinary English is manifested in the face in the same way across cultures. If a facial signal were instead a non-optional part of a prescriptive definition, like passing the Turing test is a non-optional part of a prescriptive definition of intelligence, then an event would have to show the predicted signal in order to be a basic emotion. It would not be a surprising discovery that basic emotion have distinctive signals, because something could not be a basic emotion if it did not have such signals.

So, my final query to basic emotion theorists writ large is whether they offer Ekman’s list as a definition of
“emotion,” and, if so, whether the definition is descriptive or prescriptive. If it is descriptive, where’s the evidence that captures the folk concept of emotion better than competing accounts? If it is prescriptive, how can we deal with the indeterminacy and heterogeneity implied, and how can we make the concept more scientifically measurable? Is emotion to be thought of as the package of components or as the neural mechanism underlying the changes of components? If the list instead is a set of empirical claims, what definition of basic emotion does it presuppose?

But what is the alternative? My suggestion is to treat “emotion,” “anger,” “fear,” and so on as the folk concepts they are and to continue to seek descriptive definitions for these terms. Independently, we can develop an alternative conceptual framework for all the events those terms (vaguely) point to. We would abandon the assumption that “emotion,” “anger,” and the rest pick out scientifically tractable domains and the assumption that “emotion” is an explanation.

Those events consist of components, and the components provide one (but not the only) starting point. This is precisely the insight at the foundation of psychological constructionism, which abandons the search for a prescriptive definition of emotion and the assumption that emotion is a scientifically tractable domain. What we need instead is to develop prescriptive definitions for each of the components of an emotional event, whether each occurs during an emotion or not.

This captures what experimentalists are already doing. Although an experiment might be dubbed a study of, say, fear, the experiment will in fact be generally a study of the components and only some of them. (For example, Levenson, Ekman, Heider, & Friesen’s, 1992, famous study of emotion among the Minangkabau was phrased as a study of emotion but was in fact a study of two components: specific changes in facial behavior and specific changes in the autonomic nervous system.) LeDoux (2015) has recently articulated his reasons for re-framing his own theory and research not in terms of fear but in terms of certain neural pathways that are part of a survival system. He wrote: “I have to share some of the blame for being viewed as a basic emotions kind of researcher. In my work on Pavlovian fear conditioning, I have focused on how the brain detects and responds to learned threats. But I called the brain system that does this the fear system. This was a mistake.”

K-C: The multiple criteria approach to defining emotion so central to BET is at its heart pragmatic, oriented toward empirical science. It helps researchers differentiate moods, which are longer, from emotions, which are briefer (although this claim has generated very little empirical attention). It gives researchers some solid, well-defined, measurable criteria for differentiating emotions from one another (e.g., how is embarrassment different from shame, or awe from surprise and fear). It allows researchers to identify emotions as they occur in the complex unfolding of social behavior. The peer-reviewed studies inspired by BET number in the thousands. It’s naïve to assume that such a robust empirical tradition with so many field-shaping advances represents a failure.

More generally, theoretical claims should be judged by the scientific studies they inspire. And in this sense, the criteria posited by BET have proven to be enormously useful in three distinct ways. First, these criteria of BET have proven critical to the study of emotions ignored in early studies of emotional expression, paving the way for new discoveries, a hallmark of a useful scientific approach. Guided by BET’s criteria, scientists have made significant progress in understanding emotions like embarrassment, pride, and shame, as well as sympathy, love and desire, and more recently awe, relief, and interest. Critical to these more focal endeavors is the notion to compare closely related states to one another to identify distinctions in their patterns, a framework laid out in BET.

Second, BET provides a set of guidelines for synthetic and systematic empirical review: if one wants to make the
case that a state is an emotion, the criteria spell out what sort of data one must provide (e.g. data on distinctive signals, data on distinctive physiology, etc.), and these criteria have been used with clarity and success in such reviews (e.g., on embarrassment, see Keltner & Buswell, 1997; on compassion, see Goetz, Simon-Thomas, & Keltner, 2010; on pride, Tracy & Robins, 2004). What this has encouraged is a more synthetic approach to theorizing about emotion, one that grapples with the evolution of different emotions, their likely physiological underpinnings, and the specific behaviors emotions guide.

Finally, it is worth noting that BET laid a foundation for functional approaches to emotions, which posit that emotions serve more specific proximal functions within social interactions, and more distal functions within evolutionary or cultural frameworks. This more explicit theorizing about the functions of distinct emotions is an extension of the early claims in BET that emotions are adaptations to the fundamental life tasks humans have faced in our evolution, and has proven fruitful in guiding the field.

Before turning to what we think is in need of revision, we also note a critical emphasis in this multiple criteria approach to emotion: it moves the question of what emotions are and how we measure them beyond the reliance upon words, an emphasis that dominates many areas of inquiry (e.g., studies of self-reports of emotional experience, narrative studies, emotion recognition studies of static photos). It does so because subjective experience is just one of the criteria for the instantiation of an emotion, and a non-essential one. We believe the sole reliance upon self-report measures is problematic because a) reports of emotion are retrospective, b) those reports are filtered through a judgmental context, and c) emotion words are difficult to translate across cultures. The BET definition above prioritizes signaling behavior, and physiology, which are less subject to the aforementioned problems.

But important modifications are needed. First, systematic attention must be paid to contextual sources of variation in the production of emotional behavior and the meaning of that behavior to outside observers – a vibrant area of inquiry in new contextualist approaches to emotion recognition (Barrett, Mesquita, & Gendron, 2011). This line of work will have profound implications for the study of expressive behavior, revealing individual, cultural, and contextual sources of variation in signaling behavior and away from the idea that each and every instance of an emotion is signaled in the same behavior across individuals and contexts.

Second, we note that the BET criteria do not capture how emotions unfold, and this neglects the behaviors emotions motivate and the specific effects emotions have on different cognitive processes, such as attention, perception, memory, and judgment and decision making. This is a critical shortcoming, one we hope that will be remedied as researchers move to understand how emotions unfold dynamically, shaping patterns of relating to the environment over time.

Finally, it will be important for this approach to grapple with the idea that emotions cluster in families. In its evolution, BET has moved to the position that all of its criteria apply similarly to all emotions. This is an important stance, for it clearly defines what differentiates the emotions from one another, and enables the study of new emotions, as outlined above. But, ultimately, we think the story will be more complex, with some criteria applying more directly to certain emotions than others. We have already outlined how this is so in the realm of expression: some emotions are clearly signaled in the face, others not.

It is likely that other response systems may be common to clusters of emotion for important theoretical reasons. For example, empirical studies can differentiate prosocial emotions like love, sympathy, and gratitude in terms of appraisal patterns and thoughts and tactile and vocal expression, but they may all share a core peripheral
physiological response – elevated vagus nerve activation, perhaps the release of oxytocin – but be differentiated in specific central nervous system patterns. A robust definition of emotion must illuminate not only distinctions between emotions but commonalities across conceptually similar emotions, and this is lacking in current formulations of BET. The promise of this work is a fuller understanding of how the myriad response systems now studied in the field of emotion map to the 20 or so states of interest to investigators.

AS: Thanks Dacher, Daniel, Alan and Jim for this terrific debate. For decades, research programs in the study of facial expressions have been at odds with another without directly engaging in a formal debate. You deserve great credit for having done so in this setting. It is my hope that this debate has made clearer than ever before what competing research programs in the study of facial expressions still disagree about, while also revealing areas of substantive agreement. I encourage our many readers, who, I am sure, have been excited, incensed or otherwise provoked by various aspects of this debate, to chime in, using the “Leave a comment” tool below. Thanks again!

References


emotion, 3(2), 81-123.


Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between emotion and facial expression:
Evidence from laboratory experiments. *Emotion Review, 5*(1), 16-23.


