

# From interpretation to identification: a history of facial images in the sciences of emotion

JOHN McCLAIN WATSON

## ABSTRACT

Although images of faces have long been employed in the scientific study of emotion, the objectives and assumptions motivating their use have shifted according to the various fields and research programs within which they have been put to use. This article traces these shifts through three such fields – the social psychology of interwar America, cross-cultural research of the 1970s, and the contemporary neurosciences of emotion – in order to assess the recent use of facial images as a means of correlating particular emotions with particular locations in the brain.

*Key words* emotion, facial expression, Facial Action Coding System, localization, neuroscience

Your face is a fiction. Having never seen it yourself, you take the mirror and photograph at their word, knowing that – at best – they permit only a second impression. You cringe when viewing pictures of yourself, not because you are vain, but because doing so forces you to acknowledge the chasm between your self and your surface. Estrangement from your body's most personal marker is the condition of possibility for self-other awareness. It compels you

to create technologies of recuperation – fashion, gesture, stories – with which to gloss the certainties that your face *has already put into motion*. This estrangement is the wellspring of human (mis)communication.

Not so for contemporary neuroscientists interested in the neural correlates of emotion. Utilizing brain-imaging techniques, the relation between faces and emotion is assessed in order to identify neurological structures activated when subjects are presented with computer-generated facial images. Treated as stable referents for discreet emotions, these images of faces are used unproblematically as simple prompts designed to elicit the desired pattern of brain activation. Localizing a universal human trait within a particular brain structure allows contemporary researchers to occupy both ‘universal’ and ‘particular’ topoi by accounting for both valences of the trump term *natural*; it unites ‘what is’ with ‘what must be’. The ensuing rhetoric of mastery regarding emotion contrasts markedly with the earlier approaches to the subject, and the purpose of this article is to illuminate the historico-discursive bases of that contemporary rhetoric.

In the first section of the article I discuss research papers from the 1920s and 1930s which document investigations of the relationship between facial expression and emotion. I consider the stated objectives that motivated the studies, the methods through which the images of faces used in the study were composed, and how these objectives and images are framed to produce data from which claims can be made regarding the relationship between facial expression and emotion. These older studies illustrate how early researchers struggled both with conceptual and methodological issues pertaining to the experiential and context-dependent dimension of facial expression and emotion. These issues, regarding the experimental value of affective disposition, personal history, and other situational variables that influenced interpretation of a facial expression, were considered to be elements critical for any scientific understanding of emotion.

Next, I identify what is commonly accepted as the most compelling justification for omitting these variables – and the conceptual commitments they represent – from current neuroscientific research on emotion. The Facial Action Coding System (FACS) was developed in the 1970s as a standardized coding scheme intended to enable psychologists to identify and construct facial expressions that have been shown to exhibit high cross-cultural agreement as to the emotion being presented. Over the past decade, FACS has been deployed throughout the brain sciences to such a degree that it currently helps to underwrite an entire research program designed to localize emotion to particular structures of the human brain. I argue that FACS has been positioned as a certain kind of ‘breakthrough’ in research on the relationship between facial expression and emotion because it enables brain scientists to bracket the difficult questions about emotion that were considered to be important by pioneers in the field.

To illustrate this contemporary experimental use of facial images, I present three recent examples of brain research in order to illustrate more precisely how contemporary scientists employ FACS in their research. These researchers use images analyzed or produced using FACS because they are thought to possess sufficient stability and reliability to ground investigations into and claims regarding the localizability of particular emotions within particular structures of the brain. In the end, I argue that this stability – which is crucial to the localization-of-emotion thesis – is authentic only if one is willing to bracket the important issues that arise when validating neuroscientific claims using social psychological constructs. In an era marked by the increasing number and scope of collaborations between psychology and neuroscience, the question of whether or not researchers are willing to engage in this kind of bracketing – or are even aware that it is occurring – is of increasing importance.

### ACT ONE: PERFORMING EMOTION/INTERPRETING THE FACE

The first thing that strikes one looking at facial expression research from the first half of the 20th century is the relatively limited – or, to put it another way, ‘open’ – set of objectives that motivated researchers. The rationale for Columbia psychologist Antoinette Feleky is typical: ‘It is the purpose of this article to illustrate the expressive movements characteristic of certain emotional states, or rather, to show what emotional states certain facial expressions do signify’ (Feleky, 1914: 33). In her study, volunteers, or ‘one hundred reliable persons’, were shown 86 photographs of a female posed with particular facial expressions, given a list of approximately 100 ‘names of emotions’, and told to write down ‘the name of the expression which the photograph suggests to you’ (Feleky, 1914: 34). The fact that subjects in the Feleky study were then asked to introspect and write their thoughts as they were considering how best to understand the image will become important later in the article, where it will be argued that this space for narrative invention entails the assumption that what subjects were being asked to do was *interpret* the images, not identify the emotion contained ‘within’ the image.

Twenty-four photographs of faces were included in the text of the Feleky paper and it is necessary to describe how these pictures were produced. The pictures used in the Feleky experiment, which were later used widely by other researchers, showed an individual who ‘had clearly in mind what she was endeavoring to portray, either by deliberately calling up the emotion itself, or by reciting words expressing the desired emotion’ (Feleky, 1914: 33; see Plate 1). Importantly, the stimuli meant to evoke several of the expressions are included in the article and they treat emotion as an inherently social



*Plate 1* One of the photographs from the Feleky paper (1914): an individual who ‘had clearly in mind what she was endeavoring to portray’.

and highly context-specific behavior. That is, a behavior had to be performed in order for an emotion to be produced. For example, posing for the picture that was to suggest ‘modesty’, the female student read aloud the second line in Gretchen’s speech to Faust, ‘I feel it, you but spare my ignorance. To shame me, sir, you stoop thus low.’ The image for ‘attention’ was taken when the assistant was posed ‘for attention to a purely intellectual manner. While mentally multiplying  $19 \times 19$  the subject was photographed.’

Finally, the image for ‘fear’ was ‘made after the subject had said the word “poison” in reciting Juliet’s speech in the potion scene – “What if it be a poison, which the friar subtly hath ministered, to have me dead?”’ (Feleky, 1914: 35–6). Emotion here is produced not simply through experimental instruction – ‘posed for fear’ – but via the affective engagement with and stimulation by great literature.

Feleky was quite confident in the use-value of her experimental method. Not only does the chart which presents the raw data of the investigation constitute a model through which one can examine the relative pliability of facial expression as a sign of a given emotion, but the results yield ‘in cases where the facial expressions are clearly significant, means of studying emotional expressions and illustrating them before classes in psychology or dramatic art’ (Feleky, 1914: 35). It is interesting to note that only those pictures on which there was a high percentage of interpretive agreement are offered as

productive for further study. One might just as well have argued that those images on which there were *disagreements* in interpretation would be just as important to study since by interrogating such disagreements one might be able to produce further insight into the sociality of facial expression and emotion. Nevertheless, as we will see, the Feleky experiment would not be the last time that high agreement was taken to be suggestive for further research.

The Feleky study served as a baseline for later investigations of facial expression and emotion. Whereas its primary objective was to show how people attributed emotions to various facial expressions, subsequent studies altered aspects of Feleky's experimental protocol in order to get a clearer sense for the variables that needed to be taken into consideration by researchers when investigating facial expression. That is, not only did researchers want to know how people attributed emotion, they also wanted to know what factors influenced the attribution of one emotion rather than another and why this was the case. For example, University of Iowa psychologist C. A. Ruckmick adds an important experimental protocol when he asked observers to 'describe their emotional state' before they were presented with the images of faces (Ruckmick, 1921: 32). He hypothesized that the attribution of emotion from a posed photograph had something to do with the day-to-day mood of the person being asked to view the photos and, as expected, 'It appeared that the mood of the observer changing from day to day had some effect on the interpretation of the face' (Ruckmick, 1921: 34).

The pictures for the Ruckmick study were produced in a manner similar to the Feleky pictures. Ruckmick procured the services of 'one of the talented women students in the University who had had considerable training in dramatic performances' who 'during her spare time in the morning . . . would practice the expression before a mirror, frequently with the assistance of some classical quotation which she had recalled in connection with this particular emotion' (Ruckmick, 1921: 31; see Plate 2). This student, Miss Merle Turner, is thanked in the first footnote of the article for 'eagerly taking part in the long process of making, selecting, and often discarding the expressions obtained'. In contrast to the Feleky pictures, after these photographs were taken, Turner and Ruckmick would 'repair to the dark room to see whether the negative was satisfactory. If unsuccessful, that particular expression was repeated until it had been properly reproduced' (Ruckmick, 1921: 32).

Observers in this study were given the following instructions, 'You will be shown the photograph of a face. Please note first what meaning you read into the face, and second, any change in your own affective reaction as a result of viewing the picture' (Ruckmick, 1921: 32). After the collecting of this data a chart was constructed to show the word that each observer attributed to the face in the photograph. Again, the open-ended response format yielded quite interesting interpretations. Responding to the image for 'grief', observers



*Plate 2* A picture from the Ruckmick study (1921): the student ‘would practice the expression before a mirror’.

wrote ‘dejected anxiety’, ‘questioning’, and ‘plea for mercy’. For ‘hope’, some responses were ‘longing’, ‘dreaming’, and ‘pity’. Ruckmick explains this variability by suggesting that ‘some allowance must be made for inadequacy of verbal expression, as when an observer gropes about to find a word that will carry his meaning’ (Ruckmick, 1921: 33). Then, regarding the response to how subjects’ own affective state changed after having viewed each picture, Ruckmick found that, in order to interpret the emotion of each face, observers tended to imagine a specific situation or context in which the person in the photograph might reasonably have made that particular face. Subjects might remember personal situations in which they made a similar expression, or conjure a dramatic scene through which the face in the photo could be made intelligible.

Subjects were asked to describe these mini-dramas and, even though several examples of these are included in the study, it is not noted in the article which picture motivated which description. Still, one can only guess what sort of facial expressions might have motivated the following responses (Ruckmick, 1921: 34):

She has been crying, but when another person enters the room she becomes defiant. She refuses to be dictated to. Hitherto she has been quiet but on the entrance of the second person she calls out angrily.

The rather starry eyes give the idea of interest but of a person who is ill

and has not physical strength or else is tired. I remember feeling the same way the subject looked when I have been tired.

Someone for whom the girl feels responsible has gotten into difficulties and the subject of the picture is more or less terrified and searching anxiously for a solution of the problem.

Terribly unpleasant visual perception of a drunken man and his children in the room.

The production of an emotional context from the pictures suggested, and a later experiment confirmed, that the mood of the observer was a key variable in the attribution of emotion from facial expression. Later experiments involved showing photographs of starving Viennese children to four observers or scolding them 'in the presence of others' before asking them to view the posed photographs. Under both of these conditions it was found that 'individuals vary not only with regard to suggestions from without but also in accordance with emotional experiences often concealed from another in interpreting the same facial expression on different occasions' (Ruckmick, 1921: 35).

Whereas Feleky illustrated the variability of emotional attribution based on particular facial expressions, Ruckmick's concern was to illustrate the influence that the particular mood and emotional disposition of each observer had on how he or she interpreted the picture. On the subject side, Ruckmick's inclusion of self-assessment and extended response components in the study indicates important assumptions about the impact of personal experience, experimental context, the variability of language, and social expectations in the attribution of emotion. On the investigator side, recall that Ruckmick and Turner deliberated about the veracity of each picture before it was included in the study. Just as subjects' experience and disposition affected their interpretation of the images, we can be sure that the dramatic training and sensibilities of Ruckmick and Turner likewise heavily conditioned their judgement of 'satisfactory' or 'unsuccessful' images. Furthermore, just as Turner employed literature and other dramatic arts when performing in front of Ruckmick's camera, subjects themselves exploited a wide array of narrative and cultural resources when they were asked to interpret the photographs. Even though they were only shown posed photographs of a face against a black background, both subjects and investigators found it necessary to compose an embodied social setting within which those faces could become intelligible.

It is important to point out that the interpretive and imaginative work required of Ruckmick's subject was intelligible only because allowances for that work were deliberately included in the study's experimental design. If subjects are asked to write down a couple of words, as in the Feleky study, or even several sentences, as in the Ruckmick study, through which they are

allowed to construct an interpretation of the face in the photograph, then subjects will do just that and formally valid conclusions can be produced from that data. In this way, experimental protocol defines the limits of intelligibility of a given emotion. Protocols which do not include a component wherein subjects can script an interpretive context for the image, or that otherwise neglect the interpretive over and against the identificatory nature of the subject's task, risk adversely affecting the quality but not necessarily the validity of the data from which any conclusions can be drawn.

In 1931, this concern with quality was extensively addressed in a lengthy paper by Johns Hopkins psychiatrist Leo Kanner. Kanner's paper is important here because it demonstrates in great detail how the methodological and conceptual challenges inherent when investigating the relationship between facial expression and emotion were faced head-on in the early days of facial expression research. The following excerpt gives a clear sense of Kanner's attitude regarding the methodological and moral seriousness incumbent upon those who would endeavor to understand the relationship between facial expression, emotion, and human being.

The vocabularies, phraseologies, and adages, the artistic, belletristic, and pseudo-scientific productions of all nations and ages contain almost innumerable references to conclusions drawn from an individual's facial expression as to his emotional state, intellectual endowment, and ethical qualities. Anthropomorphized deity, made to combine the ideal perfections of all desirable human properties, is credited in practically all religious systems with an unlimited capacity to read the thoughts and emotions of man like an open book. In the day dreams of mankind, in the fantasy life of many persons, in the quack utilization of the unsatisfied and not really satisfiable wishes of an untutored laity, in the autistic schizophrenic delusions removed from the need of any checking consideration of actual possibilities, the idea of mind-reading has always been one of the consorts of the equally mythologic 'phenomena' of telepathy, spiritualism, horoscope prophecies, and fortune telling. (Kanner, 1931: 1)

Clearly, the reader is being prepared for a break with this sad history. Rather than update the aporias produced when attempting to deduce emotion, ethics, and intellect from facial expression, Kanner seeks to redeem the face as an object of study by rigorously accounting for each of the major social, psychological, and environmental factors that influenced: (1) the facial expression of emotion; (2) the ways emotion gets attributed to facial expression; and (3) how scientific investigators have and should go about assessing the relationship between (1) and (2).

Kanner's protocol was fairly straightforward: students were presented with a dozen Feleky photos and asked to interpret each picture by 'choosing the

best term they could find which would do full justice to the type or shade of emotional reaction'. Observers were also given a space in which to write the 'concrete situation, which, in the observer's opinion, might in each instance have given rise to the interpreted expression' (Kanner, 1931: 12). Kanner employed this design because it allowed him to demonstrate and problematize core issues underlying all aspects of the attempt to assess the relationship between facial expression and emotion. The first challenge regarded the practice of giving subjects a list of emotional labels from which they must choose as they attempted to assess the photographs. For several reasons having to do with (1) undue suggestion; (2) the curtailment of observers' 'linguistic inventiveness'; (3) forced choice; (4) the fact that 'use of a term does not exclude ignorance of its accepted meaning'; and (5) the inability to accommodate a realistic amount of combination or difference in adjectives, Kanner decided to abandon the prescribed list of emotional labels and instead had students 'depend on their own vocabularies' (Kanner, 1931: 13). Subjects labeled the images freely without choosing terms from a predetermined list. This small experimental modification had tremendous conceptual implications. It meant that attributions could not be judged 'correct' or 'incorrect' since there was no standard or 'correct' set of interpretations from which to make such a judgement. Furthermore, the absence of any correct/incorrect baseline of judgement meant that the ultimate validity of the conclusions reached in the experiment would only be as good as the analysis and interpretation performed by the investigators as they went about assessing the data.

After gathering subject responses, Kanner commenced a thoroughgoing critical assessment of method that he divided into four headings: (1) Quality of the Presented Photographs; (2) the Linguistic Aspect of the Experiment; (3) the Situational Factor; and (4) the Personality Aspect. Feleky's 1914 suggestion that future researchers use only those photographs on which there was a high degree of agreement across observers was taken by Kanner, who takes the quality of the Feleky photographs more or less for granted. In any event, the quality of the picture becomes less important once Kanner has agreed to take subjects' descriptions at face value. Kanner then critiqued what he called the 'linguistic aspect' of the experiment, extensively analyzing each semantic, etymological and symbolic dimension of selected responses from subjects. Underlying this linguistic dimension was the fact that 'the names of emotional expressions, emotional attitudes, emotional reactions in general are not names for any existing entities. . . . There is a danger that these forms of diction be taken by those who use them as full truths existing *per se* instead of as linguistic symbolizations' (Kanner, 1931: 24). In other words, Kanner argued that researchers tended to take operationalization too far by conflating the emotional label with the emotion 'itself'.

Of course, the notion that there is an uncertain relationship between an emotion and its linguistic label is not new. To take perhaps the most obvious

example, the critique of introspection as a valid psychological method was based in part on an assumption that language-use was hopelessly incapable of transmitting the nature of psychological, much less physiological, states (Uttal, 2000: 93). This assumption forced a kind of choice between differing hierarchies of value, and it is in the wake of this choice that we frame much of the so-called 'problem of language'. Within the contemporary context, marked by renewed interest in the more direct impact of biology on behavior, the diffusion of brain-imaging technologies, and the proliferation of alliances between and across various research programs in psychology and neuroscience, the arbitrariness of language gets framed as a liability which must be overcome through the development of techniques that can bypass or veil its deleterious effect. Since brain activity is believed to be untainted by the sociality and contingency of all things verbal, it is believed that attending to the brain and the brain only will produce more reliable knowledge by going straight to the (alleged) source of emotion.

In the section discussing the 'Situational Factor', Kanner makes clear his position on the importance of context and interpretation in the study of the relationship between facial expression and emotion. 'Any judgment of emotional expression remains without much significance if it is not viewed as the recognition of an indicator of something that is happening, or that has happened, or that is expected to happen' (Kanner, 1931: 33). Curiously, though, after a few examples to illustrate his point, the subject of the sociality of emotion and facial expression is dropped. Kanner's final critique regards the 'Personality Aspect' of facial expression research, which he divided into discussions regarding the personality of the investigator, of the subject, and of the person expressing the emotion. First, Kanner confronts criticism that the scoring of subject responses is due as much to the personality of the scorer as it is to the interpretation under investigation. His response must be quoted in full:

Is not the pluralistic attitude, illustrated in this paper perhaps a little too sketchily, an attempt to include all important factors entering into the problem, which saves one from one-sidedness and arbitrariness and from doing gross injustice to the object under study? It is very true that one engaged in any scientific research in the field of human activities must bring with him a certain familiarity with the facts and an ability to single out those topics which are relevant and indispensable for the objectivity of the results. If one keeps this in mind, it will, considering certain natural limitations inherent in all psychobiological work, not matter so much if one becomes guilty of an occasional deviation of one or two points from what another individual, working with the same facts, might obtain. The various criteria for grading discussed above warrant as much precision as is possible in the observation of human beings. (Kanner, 1931: 35)

In other words, some differences in valuation will occur but it is the responsibility of the researcher to consider every possible variable in the attempt to understand this most significant human activity. For Kanner, it seems, anything less would be immoral. Regarding the personality of the subjects, Kanner suggests not only that one can never be sure that someone *else* is having an emotion, but even that the person putatively experiencing the emotion may not know it, much less be able to communicate it accurately. Finally, regarding the personality of the person expressing the emotion, Kanner reminds us that, unlike the photos used in facial expression research, facial expressions outside the laboratory are rarely intended for recognition. That is, 'actual' facial expressions are much more difficult to pin down and so, to some extent, posed photographs oversimplify the task of interpreting a given image of a face.

To summarize Kanner's evaluation, then, he illustrates (1) problems with providing subjects with a list of emotions to choose from and problems with *not* giving them a list to choose from; (2) problems with investigating the synonymic and etymological histories of the attributed emotional terms and problems with *not* investigating the use history of emotional terms; and (3) problems with using posed photographs and problems with *not* using posed photographs. The challenges are even more evident when it comes to evaluating the data once it has been generated. Indeed, remarking on ultimate hopes of scientifically investigating the relationship between emotion and facial expression, Kanner admits 'A comparison of these scales may tend to make one unnecessarily pessimistic as to the value of the entire investigation' (Kanner, 1931: 39). And yet, despite his incisive critique, Kanner remained steadfastly committed to an objective of 'building up a broad and workable basis for additional research' (Kanner, 1931: 51).

I go into such detail about Kanner's work in order to show how seriously the methodological and conceptual difficulties inherent in assessing the relationship between facial expression and emotion were taken by early 20th-century psychologists. The 1920s and 1930s were, of course, marked by a transition away from the more speculative and descriptive approaches of 19th-century psychology and towards experimental and reductionistic approaches through which basic psychological processes could be deduced through self-report, observational and, later, physiological techniques. It would therefore be misleading to suggest that Kanner did not at some level want to produce generalizable, perhaps even propositional, statements about the relationship between facial expression and emotion. Nevertheless, it is enough that the difficulties associated with this kind of research were confronted head-on and considered necessary to resolve before the relationship between facial expression and emotion could be adequately understood.

INTERLUDE: UNIVERSALIZING AND  
BIOLOGIZING THE FACE

The remainder of this article investigates some of the ways that the relationship between facial expression and emotion is being pursued in the contemporary neurosciences. Although there are both affinities and differences between these more recent studies and those conducted by early 20th-century investigators, in what follows I argue that the now dominant neural perspective on emotion and facial expression serves to *bracket* rather than resolve the conceptual and methodological issues that were so vexing to early investigators. Granted, a rigorous appreciation of the myriad variables, desires and circumstances that inform and condition emotional display and emotional attribution – variables once considered necessary for a valid and rich understanding of facial expression and emotion – may be less important in the context of behavioral neuroscience than it might be in the context of social psychology. And yet the case of facial expression research is a very clear illustration of how those two contexts are (and have always been) much more complicated than we tend to let on. In the present case, it turns out that the reason neuroscientists are able to bracket questions of language, context, and experience is that they utilize in their experiments a set of facial images which was created, theorized, and validated squarely within the context of social psychology. The practical and conceptual issues considered of utmost importance in these images' *context of origination* do not travel with them when they get picked up and put into the service of knowledge creation in a different *context of application*. And yet, while the kind of knowledge being produced via pictures of faces in recent neuroscientific research is of a decidedly reductionist and physiological kind, it is by no means clear that these images – or the system by which they are understood – provide grounds that are as unambiguous as the claims being made upon them. The question to be answered is: what happened after Ruckmick that made it possible for neuroscientists to rely on images of faces as an instrument for making (apparently) determinate claims about the locations of emotion in the brain?

Answering this question requires a brief detour through social psychological research on facial expression in the 1960s and 1970s. It was during this time that an auspicious union was forged between Tompkins's (1962) psychology of affect and Hjortsjo's (1970) pioneering anatomical investigation of the visible appearance changes of each facial muscle. This union was accomplished by a group of psychologists who set out to construct a comprehensive, descriptive system with which expression could be analyzed according to minimal units of facial behavior. Earlier facial expression coding schemes had been devised in order to answer experiment-specific questions about the relationship between facial behavior and emotion, but a comprehensive system – and the common nomenclature that it would introduce –



*Plate 3* An Ekman *et al.* (1969) image; cross-cultural similarities suggested a 'pan-cultural' nature of facial expression.

was necessary if the scientific study of facial expression was to advance (Ekman and Friesen, 1978).<sup>1</sup> But which expressions were to serve as the prototypes upon which this comprehensive system should be grounded? The

answer turned out to be easy, thanks to psychologists Paul Ekman, Richard Sorenson, and William Friesen, who had just completed a set of highly influential cross-cultural studies providing intriguing evidence for the universality of certain facial expressions. Using images of variously posed Caucasian faces, Ekman *et al.* (1969) conducted multiple experiments (in Japan, New Guinea, Borneo, Brazil, and the United States) that sought to determine whether or not subjects tended to attribute more or less the same emotional label to the same image. Despite widely differing levels of literacy, exposure to media, and geography, it was found that a handful of these images achieved between 70 per cent and 80 per cent agreement when subjects were asked to label them using one of six emotional labels: anger, disgust, fear, joy, sadness, and surprise (Ekman *et al.*, 1969; see Plate 3). Follow-up studies further validated the 'pan-cultural' nature of facial expression (Ekman, 1971; Ekman, Friesen and Ellsworth, 1972; but see Russell, 1994, 1995, and Fridlund, 1994).

It may be somewhat surprising to learn that not until the 1970s was a system for measuring and analyzing facial expression developed. One would have thought that such coding schemes would have been produced as a matter of course within the increasingly technologized field of postwar experimental psychology. And yet the empirical validity and conceptual stability necessary for such a system simply did not exist until the neurocultural theory was formally introduced by Ekman in 1972. In this way, the high-agreement images that sanction the neurocultural theory of facial expression came to acquire a wholly unique conceptual value in the sciences of emotion. Not only were they interpreted similarly by most humans but, because of this, they were seen to possess precisely the prototypic quality considered necessary for the construction of a comprehensive, anatomically based facial coding scheme. In 1976, Ekman and Friesen made these high-agreement images available in slide form under the title *Pictures of Facial Affect* and in 1978 they published a coding scheme based on those images in a monograph titled *The Facial Action Coding System* (FACS).<sup>2</sup> FACS is said to provide a

... comprehensive, anatomically based system for measuring all visually discernible facial movement ... FACS describes all visually distinguishable facial activity on the basis of 44 unique action units (AUs), as well as several categories of head and eye positions and movements. Each AU has a numeric code, the designation of which is fairly arbitrary. ... FACS coding procedures also allow for coding of the intensity of each facial action on a 5-point intensity scale, for the timing of facial actions, and for the coding of facial expressions in terms of events. (Ekman and Rosenberg, 1997: 12–13)

In short, FACS is said to provide a standard baseline against which researchers can test their hypotheses regarding the relationship between facial

expression and emotion. Over many years and hundreds of experiments, images created or coded using FACS have been used to identify very specific muscular behaviors which, when combined, have led a statistically significant number of people to interpret a posed expression in more or less the same way. These and similar images have been taken up in the neurosciences of emotion as evidence of the phylogenetic bases of particular facial expressions which must, of course, have a neurophysiological origin. FACS enables researchers both to code already existing images of faces and to create novel expressions by manipulating images of faces on a computer. Each of the neurological studies described below employs Ekman pictures and treats them as evidence of a species-universal display of emotion.

Of course, as Griffiths points out, while Ekman's studies 'show that people in all cultures respond in a similar way to things that frighten them', 'they do not show that people in all cultures are frightened of the same things' (Griffiths, 1997: 55). Nevertheless, and despite Kanner's early concerns about the conceptual difficulty of judging a subject's attribution of emotion as 'correct' or 'incorrect', the use of Ekman's images gives contemporary brain researchers a standard against which they can judge subjects' interpretations as correct or incorrect. This in turn authorizes them to make claims about the role that subjects' damaged neural structures might play in causing them to mis-identify a presented image. These images are treated in neurological research as stable referents for emotion, and it is this (alleged) certainty that sanctions neuroscientific claims to have localized, for example, 'fear' to the amygdala (Adolphs, Tranel and Damasio, 1998: 472). The concerns of Kanner, Ruckmick and Feleky, who struggled to attend to the role of mood, context, language, and other factors central to the practice of interpretation, can be overlooked thanks to the creation of images that are (alleged to be) immune from those confounding variables.

## ACT TWO: SECURING (THE) FACE IN THE AGE OF NEUROSCIENCE

In the following pages I present brief snapshots of three neuroscientific research papers whose knowledge contribution can be attributed directly to FACS. That is, each of these studies presumes that the conceptual and empirical validity of FACS has already been vetted and that its primary images and the decoding scheme by which they are to be understood can be used as innocently as one would use a thermometer to measure temperature. Indeed, if one way to determine the status of intellectual tools like FACS is by assessing the citational facility with which they are employed, then FACS and the images in *Pictures of Facial Affect* (1976) are well on their way to possessing unquestioned authority within the neurosciences of emotion. It is exceedingly rare

that the history or legitimating assumptions of FACS are mentioned in the neuroscientific studies which make use of it. Instead, images are said to be 'facial stimuli from a standard series' (Critchley *et al.*, 2000: 2205) or, more often, simply 'images of happy [or fearful or sad or angered or neutral] faces'. Many, if not most, studies simply reference *Pictures of Facial Affect* and leave it at that. Once an incredibly rich – and open – question of massively over-determined interpretation, the question of the relationship between facial expression and emotion would appear now to have been closed.

University of Iowa neuroscientist Ralph Adolphs and his coauthors envision their study as 'part of an effort to map neural systems involved in the processing of emotion' which 'focuses on the possible cortical components of the process of recognizing facial expressions' (Adolphs *et al.*, 1996: 7678). They hypothesize that 'the cortical systems most responsible for the recognition of emotional facial expressions would draw on discrete regions of right higher-order sensory cortices and that the recognition of specific emotions would depend on partially distinct system subsets of such cortical regions' (Adolphs *et al.*, 1996: 7678). This hypothesis was tested using lesion analysis in 37 subjects with focal brain damage. Lesions are generally attained through natural stroke or therapeutic neurosurgical lobotomy, and they serve to disable or 'knock out' particular sections of cortex depending on the unique circumstances of the injury/surgery. Out of 37 subjects, 22 had lesions on the right hemisphere, 13 had lesions on the left hemisphere and 2 had bilateral lesions. Subjects were

... shown black-and-white slides of faces with emotional expressions and were asked to judge the expressions with respect to several verbal labels (the adjectives that corresponded to the emotions we showed). We chose 39 facial expressions from Ekman and Friesen that had all been shown to be identified reliably by normal subjects at >80 per cent success rate. (Adolphs *et al.*, 1996: 7679)

Note how subjects were given pre-set labels from which to choose in the attribution of emotion. Correlations between subjects' ratings of expression and those of normal subjects on the six emotion labels were calculated and then mapped onto a composite image showing both the section of the brain that was damaged and the subject's score on the facial recognition task. These values were separated out for each tested emotion and used to produce an image that shows the average location of lesion for each emotion. These images were then composited in order to come up with a single image to represent the two locations which are most likely to affect impairment of the ability to interpret emotion from facial expression. In the end: (1) no impairment was found in subjects who had left hemisphere lesions, (2) all subjects correctly attributed expressions of happiness, and (3) impaired recognition of facial expressions pertained only to a few emotions, especially fear. Due to

the fact that the right hemisphere is also 'preferentially involved' in emotional expression and experience, Adolphs *et al.* 'believe that these findings support the hypothesis that'

... during development, the human infant/child acquires the connection between faces expressing fear and the conceptual knowledge of what fear is (which includes instances of the subject's experience of fear). Such a process requires two neural components: (1) a structure that can link perceptual information about the face to information about the emotion that the face denotes; and (2) structures in which conceptual knowledge of the emotion can be recorded, and from where it can be retrieved in the future. (Adolphs *et al.*, 1996: 7686)

The two locations in the brain which are believed to be 'important components of the system by which adults retrieve knowledge about facial expressions of emotion' are the right inferior parietal cortex on the lateral surface and the anterior intracalcarine cortex on the mesial surface.

The Adolphs *et al.* study illustrates how contemporary neuroscientific researchers employ FACS images as stimuli through which to elicit desired brain activity. The objective is not, as was the case in earlier research, to examine which interpretations get made nor to examine patterns of attribution across subjects, but rather to see the differences between the ways in which normal and damaged brains react to an identical stimulus. Since the images come with a pretested, 'correct' emotional label, researchers explain 'incorrect' judgements from lesioned subjects by reasoning that there is a physiological connection between the location of their lesion and the emotion 'in' the image. In this way, images produced using FACS have been put in the service of the localization project currently occupying great segments of the neuroscientific community. These images enable researchers to introduce reliability and consistency into their investigation but it is important to remember that these advantages are procured only by bracketing the conceptual quandaries so thoroughly detailed by early researchers such as Ruckmick and Kanner.

The task of localization is furthered by a study which used functional magnetic resonance imaging (fMRI) to 'examine the neural substrate for perceiving disgust expressions' (Phillips *et al.*, 1997: 495). In this study subjects viewed grey-scale FACS-generated images of faces depicting disgusted, fearful and neutral expressions (see Plate 4). Phillips *et al.*'s use of standardized images goes further than the Adolphs study in that here the images are made more or less intense by manipulating the intensity levels in accordance with FACS. These images were computer-processed in order to reflect two levels of intensity, 75 per cent for less intense and 150 per cent for more intense. The so-called 'neutral expression' image was created by scaling down a happy expression by 25 per cent.<sup>3</sup>

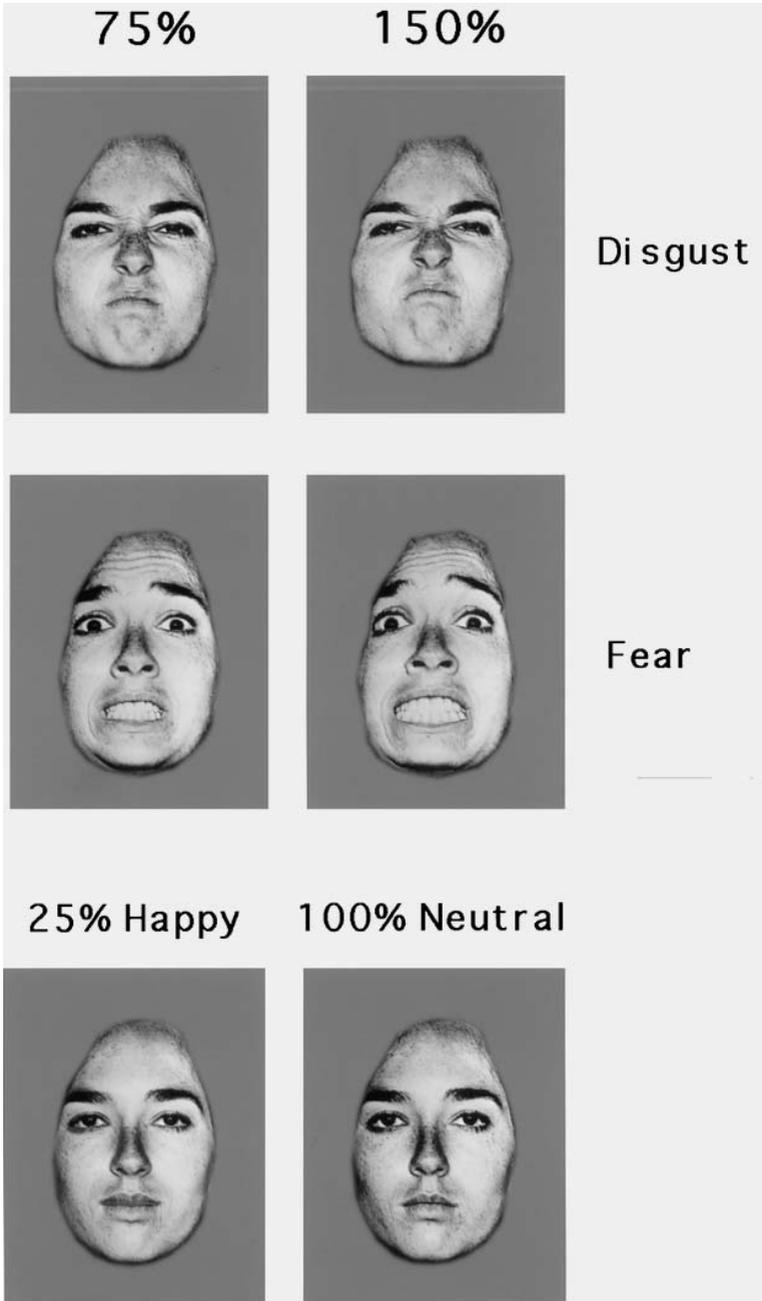


Plate 4 A FACS-generated image from the Phillips *et al.* study (1997).

Interestingly, subjects in this study were not made aware that the aim of the study was to investigate responses to emotional expression. Rather than subjects being asked to interpret the emotion presented by the facial images, subjects were shown the pictures in random order and asked to identify the sex of the person depicted in the image. 'The sex decision task was chosen to allow an identical task and response across all conditions, and to permit comparison to a previous study of fear which also used this procedure' (Phillips *et al.*, 1997: 495). By eliminating all indications that the experiment has anything to do with facial expression – and by therefore presumably ensuring that subjects will not be biasing the data by thinking about the relationship between facial expression and emotion – Phillips *et al.* illustrate perhaps most clearly how far removed questions of interpretation, context, and experience are from current neuroscientific concerns. Subjects were placed in a standard head coil in order first to measure and then reveal those parts of the brain that were activated during the sex decision task. These images 'demonstrate for the first time evidence for a differentiation between the neural responses to facial expressions of two negative emotions, fear and disgust' (Phillips *et al.*, 1997: 496).

A similar study seeking to localize emotion to a particular part of the brain claimed to have 'found that increasing activity of sad facial expression was associated with enhanced activity in the left amygdala and right temporal lobe' (Blair *et al.*, 1999: 883).

Whereas Phillips used fMRI to show correlates of neural activity, Blair *et al.* employed Positron Emission Tomography (PET) in order to show patterns of metabolic activity and thereby construct an image with greater spatial resolution. The experimental procedure was similar to Phillips *et al.*,

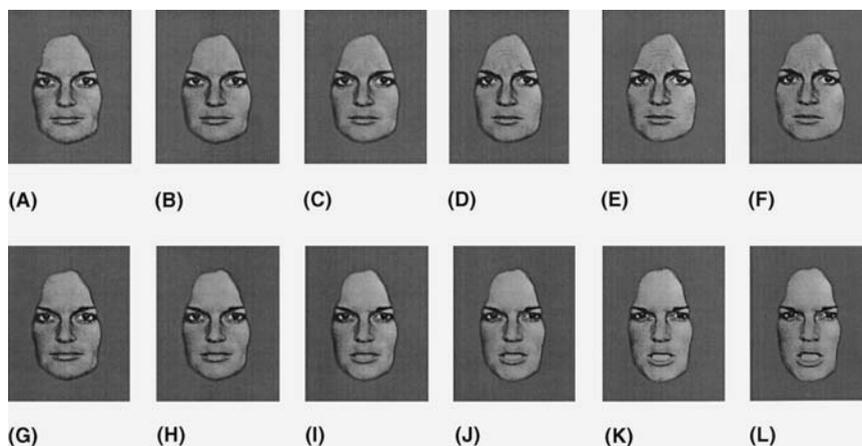


Plate 5 An image created by PET, used in Blair *et al.*'s study (1999).

with a sex discrimination task being the only direction given to subjects. Furthermore, computers were used to manipulate the intensity of expressions that were presented to subjects. 'The faces depicted either sad or angry expressions. For each emotional category and individual face, a range of six intensity levels was produced by computer graphical manipulation. The 20, 40, 60 and 80 per cent faces were interpolations created using computer morphing procedures' (Blair *et al.*, 1999: 885; see Plate 5). The authors found that the intensity of the sad expression was positively correlated with the intensity of brain activity, especially in the left amygdala, right temporal pole, right inferior temporal gyrus and right middle temporal gyrus.

### THE BEGINNING

It seems clear that the general program of localizing particular emotions within discreet areas of the human brain is only beginning. I have argued that a central prerequisite of this productivity is the escalating status of FACS and the conceptual assumptions which undergird it. To be clear, contemporary neuroscientists make technically valid claims when they localize the ability to interpret emotion within particular brain structures. These researchers make appropriate use of the conceptual and methodological tools that are accepted in their field and even, as we have seen, rely heavily on elements from other fields in order to authorize their claims about localization. This article has attempted to contextualize the nature of this authorization. By treating facial expression as a stable referent of emotion, and emotion as an effect of neural activity, the question is begged: at what point did the referent become stable and how was this determined? I have argued that the referent became fixed once a particular set of images – conceptualized and experimentally accounted for in the conceptual environment of social psychology – became validated by virtue of their mobilization in the quite different experimental and conceptual environment of the contemporary neurosciences. The mobilization of FACS throughout the brain sciences has authorized neuroscientists to bracket questions that were central to early inquiries into the relation between facial expression, emotion, and behavior in general. These questions were raised by researchers who struggled to attend to the sociality and embodied dimension of emotion and facial expression as something inextricably bound up in language, environment, and other elements of situated context. Rigorous analysis of these elements by psychologists in the social psychological and cognitive traditions has flourished since the development of the Facial Action Coding System and there is an ongoing and vigorous debate among emotion researchers about its central assumptions and methodological shortcomings (Russell and Fernandez-Dols, 1997; Sayette *et al.*, 2001).

The reluctance of contemporary neuroscientists to similarly engage in examination of the relationships between social context and emotion or between personal experience and attribution of emotion is especially noteworthy when one remembers that FACS was originally based on an empirical protocol which had subjects undergo forced choice, self-report interpretation of still photographs (Ekman, 1971). The infinite degree of manipulation afforded by FACS encourages neuroscientists to treat photographs as fixed referents of emotion and to employ them as firm grounds for the determination of neural correlates of emotion. And yet, once one takes a cursory look at the historical, conceptual, and empirical apparatus through which this stability has been achieved, it seems clear that this foundation is far less stable than it seems.

### NOTES

I wish to thank Daniel M. Gross and three anonymous reviewers for their assistance in the preparation of this manuscript.

- 1 Alternatives to FACS include: the Facial Action Scoring Technique (FAST: Ekman, Friesen and Tomkins, 1971), the System for Identifying Affect Expressions by Holistic Judgment (AFFEX: Izard, Dougherty and Hembree, 1983), EMFACS (Ekman and Friesen, 1982), Monadic Phases (Tronick, Als and Brazelton, 1980), the Maximally Discriminative Facial Movement Coding System (MAX: Izard, 1979), and FACEM (Katsikitis and Pilowsky, 1988).
- 2 FACS was essentially a revision of Hjortsjo's (1970) 'objective and unobtrusive technique useful for visible movements, based on facial anatomy' (Russell *et al.*, 2003: 336).
- 3 The authors explain how they created their 'neutral' expression thus: 'In view of the fact that 100 per cent neutral (muscles relaxed) faces from the standard set can appear slightly cold and threatening because it is conventional to signal approval in normal social interaction, we used as the neutral baseline stimulus a very slightly happy expression (75 per cent neutral, 25 per cent happy)' (Phillips *et al.*, 1997: 498).

### BIBLIOGRAPHY

- Adolphs, R., Damasio, H., Tranel, D. and Damasio, A. R. (1996) 'Cortical Systems for the Recognition of Emotion in Facial Expressions', *Journal of Neuroscience* 16: 7678–87.
- Adolphs, R., Tranel, D. and Damasio, A. R. (1998) 'The Human Amygdala in Social Judgment', *Nature* 393: 470–4.
- Blair, R. J. R., Morris, J. S., Frith, C. D., Perrett, D. I. and Dolan, R. J. (1999) 'Dissociable Neural Responses to Facial Expressions of Sadness and Anger', *Brain* 122: 883–93.

- Critchley, Hugo D., Daly, Eileen M., Bullmore, Edward T., Williams, Steven C. R., Amelsovoort, Therese Van, Robertson, Dene M., Rowe, Andrea, Phillips, Mary, McAlonan, Grainne, Howlin, Patricia, and Murphy, Declan G. M. (2000) 'The Functional Neuroanatomy of Social Behaviour: Changes in Cerebral Blood Flow when People with Autistic Disorder Process Facial Expressions', *Brain* 123: 2203–12.
- Ekman, P. (1971) 'Universals and Cultural Differences in Facial Expressions of Emotion', in J. Cole (ed.) *Nebraska Symposium on Motivation*. Lincoln: University of Nebraska Press, pp. 207–83.
- Ekman, P. and Friesen, W. V. (1975) *Unmasking the Face*. Englewood Cliffs, NJ: Prentice Hall.
- Ekman, P. and Friesen, W. V. (1976) *Pictures of Facial Affect* [slides]. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P. and Friesen, W. V. (1978) *Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P. and Friesen, W. (1982) 'Rationale and Reliability for EMFACS Coders', unpublished.
- Ekman, P. and Friesen, W. V. (2002) 'Measuring Facial Movement', in the *Facial Action Coding System Investigator's Guide*. HTML Demonstration Version. Retrieved 13 October 2003 from <http://face-and-emotion.com/dataface/facs/guide/FACSIV1.html>
- Ekman, P., Friesen, W. V. and Ellsworth, P. (1972) *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. New York: Pergamon Press.
- Ekman, P., Friesen, W. V. and Tomkins, S. S. (1971) 'Facial Affect Scoring Technique (FAST): a First Validity Study', *Semiotica* 3: 37–58.
- Ekman, P. and Rosenberg, E., eds (1997) *What the Face Reveals*. New York: Oxford University Press.
- Ekman, P., Sorensen, E. R. and Friesen, W. V. (1969) 'Pan-Cultural Elements in Facial Displays of Emotion', *Science* 164: 86–8.
- Feleky, A. M. (1914) 'The Expression of the Emotions', *Psychological Review* 21: 33–41.
- Fridlund, A. J. (1994) *Human Facial Expression*. San Diego, CA: Academic Press.
- Griffiths, P. E. (1997) *What Emotions Really Are*. Chicago, IL: University of Chicago Press.
- Hjortsjo, C. B. (1970) *Man's Face and Mimic Language*. Lund: Student-litteratur.
- Izard, C. E. (1979) 'The Maximally Discriminative Facial Movement Coding System (MAX)'. Newark: University of Delaware, Instructional Resource Center.
- Izard, C. E., Dougherty, L. M. and Hembree, E. A. (1983) 'A System for Identifying Affect Expressions by Holistic Judgments'. Unpublished manuscript, University of Delaware.
- Kanner, L. (1931) 'Judging Emotions from Facial Expression', *Psychological Monographs* 94: 3.
- Katsikitis, M. and Pilowsky, I. (1988) 'A Study of Facial Expression in Parkinson's Disease Using a Novel Microcomputer-Based Method', *Journal of Neurology, Neurosurgery and Psychiatry* 51: 362–6.

- Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrews, C., Calder, A. J., Bullmore, E. T., Perrett, D. I., Rowland, D., Williams, S. C. R., Gray, J. A. and David, A. S. (1997) 'A Specific Neural Substrate for Perceiving Facial Expressions of Disgust', *Nature* 389: 495–8.
- Ruckmick, C. A. (1921) 'A Preliminary Study of the Emotions', *Psychological Monographs* 30: 30–6.
- Russell, J. A. (1994) 'Is There Universal Recognition of Emotion from Facial Expression? A Review of the Cross-Cultural Studies', *Psychological Bulletin* 115: 102–41.
- Russell, J. A. (1995) 'Facial Expressions of Emotion: What Lies beyond Minimal Universality?', *Psychological Bulletin* 118: 379–91.
- Russell, J. A., Bachorowski, J. A. and Fernandez-Dols, J. M. (2003) 'Facial and Vocal Expressions of Emotion', *Annual Review of Psychology* 54: 329–49.
- Russell, J. A. and Fernandez-Dols, J. M. (1997) *The Psychology of Facial Expression*. Cambridge: Cambridge University Press.
- Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A. and Perrott, D. J. (2001) 'A Psychometric Evaluation of the Facial Action Coding System for Assessing Spontaneous Expression', *Journal of Nonverbal Behavior* 25: 167–86.
- Tompkins, S. S. (1962) *Affect, Imagery, Consciousness*, Vol. 1. New York: Springer Publishing.
- Tronick, E., Als, H. and Brazelton, T. B. (1980) 'Monadic Phases: a Structural Descriptive Analysis of Infant–Mother Face-to-Face Interaction', *Merrill-Palmer Quarterly of Behavior and Development* 26: 3–24.
- Uttal, W. R. (2000) *The War between Mentalism and Behaviorism*. Mahwah, NJ: Lawrence Erlbaum Associates.

### BIOGRAPHICAL NOTE

JOHN McCLAIN WATSON is a PhD candidate in the Department of Communication Studies at the University of Iowa. He is writing a dissertation which maps the circulation of developmental neuroscience throughout discourses concerning child development, parenting, pedagogy, and social welfare.

*Address:* Department of Communication Studies, 105 BCSB, University of Iowa, Iowa City, IA 52242, USA. [email: mcclain-watson@uiowa.edu]