

The Emotional Coherence Controversy: Evidence from Human and Automatic Expression Recognition

Damien Dupré
Business School
Dublin City University
Dublin, Ireland
damien.dupre@dcu.ie

Anna Tcherkassof
Laboratoire Inter-universitaire de Psychologie
Université Grenoble Alpes
Grenoble, France
anna.tcherkassof@univ-grenoble-alpes.fr

Abstract—While it has been taken for granted in the development of several automatic facial expression recognition tools, the question of the coherence between subjective feelings and facial expressions is still a subject of debate. On one hand, the behaviorist approach conceives emotions as genetically hardwired and therefore being genuinely displayed through facial expressions. On the other hand, the constructivist approach conceives emotions as socially constructed; the emotional meaning of a facial expression being inferred by the observer. In order to evaluate the coherence between the subjective feeling of emotions and their recognition based on facial expression, 232 videos of encoders recruited to carry out an emotion elicitation task were annotated by 1383 human observers as well as by an automatic facial expression classifier. Results show low accuracy of human observers and of the automatic classifier to infer the subjective feeling from the facial expressions displayed by encoders. They also show a weak coherence between self-reported emotional states and facial emotional displays. Based on these results, the hypothesis of genetically hardwired emotion genuinely displayed is difficult to support, whereas the idea of emotion and facial expression socially constructed appears to be more likely. Accordingly, automatic emotion recognition tools based on facial expressions should be questioned.

Index Terms—Facial expression, self-report, human observer, automatic recognition.

I. INTRODUCTION

With the development of commercial automatic facial expression recognition tools [1], industries and governments are gradually implementing this technology in order to track humans' emotions in various scenarios (*e.g.*, in marketing, healthcare, and the automotive industry to name a few). This technology rests on the premise that facial expressions provide a direct access to individuals' subjective feelings also called "emotional coherence". Even if this premise is central to the modern mainstream approach of human emotion, recent research in affective science is challenging it. Once the two competing approaches are briefly described along with the predictions they respectively entail, an experiment testing these hypotheses will be presented and its results analyzed in order to provide empirical evidence to contribute to answer the question.

A. The Behaviorist Approach

Based on the behaviorist approach initiated by Darwin in *The Expression of the Emotions in Man and Animals* [2],

facial expressions are conceived as a genuine displays of an individual's inner emotional state. This hypothesis is used as a basis for the Basic Emotion Theory which states that a set of six emotions are universally displayed and are genetically hardwired not only in humans [3] but also in different animal species [4]. According to the view of the Basic Emotion Theory, "*when emotions are aroused by perception of a social event, a set of central commands produce patterned emotion-specific changes in multiple systems, including [...] facial expressions.*" [5, p. 49]. To respond to criticisms, several amendments have been made to the Basic Emotion Theory, increasing the number of basic emotions from six to seven [6] as well as adding the concept of "display rules" to explain cultural differences in the management of facial expressions [7].

Even if this theory obtained popular support, it fails to explain how individuals can feel emotions without expressing them, and how individuals can express emotions without feeling them, in instances in which display rules cannot be called upon [8], [9]. These evidences have led to an alternative conception, notably the social constructivist approach.

B. The Social Constructivist Approach

Detractors of the Basic Emotion Theory consider emotion not as genetically hardwired but as a learnt association between a given situation and an appropriate response [10], [11]. For the proponents of the constructivist approach, emotions are "concepts" based on past experiences and which are "*a collection of embodied, whole brain representations that predict what is about to happen in the sensory environment, what the best action is to deal with impending events, and their consequences for allostasis*" [12, p. 12]. Following this assumption, faces are best conceived as tools displaying signals in social interactions [13]. These signals can convey individuals' motivations and readiness [14] or social messages [15]. Therefore, facial expressions are thought as behaviors which meaning is inferred by the observer. Findings support this observer dependence [16], [17]. They show that to make meaning of another person's facial behavior, the perceiver relies in particular on her/his knowledge about emotion categories.

Regarding emotional facial expression (EFE) recognition, behaviorist and constructivist approaches lead to two divergent predictions. The former postulates that, when triggered, each basic emotion is expressed by a prototypical face (non basic emotions being blends of the basic ones). Basic emotions are easily recognized by all human observers and emotional states are accessible by facial measurement. As the recognition of facial expressions is based on the identification of specific patterns of facial movements, it implies that EFE recognition of both human observers and automatic classifiers should be as accurate. The constructivist approach, in contrast, affirms that facial expressions do not provide a direct access to individuals' subjective feelings. Even if the face possibly moves during an emotional episode, facial muscle movements are not linked in a one-to-one manner to a specific discrete emotional experience [18, p. p418]. Instead, emotions are mentally constructed by the perceiver and mental categories of emotions are needed to accurately categorize facial movements. Therefore, since there is no emotional prototypical face, one should expect human observers' superior ability to accurately recognize EFE as compared to automatic EFE recognition tools (Figure 1).

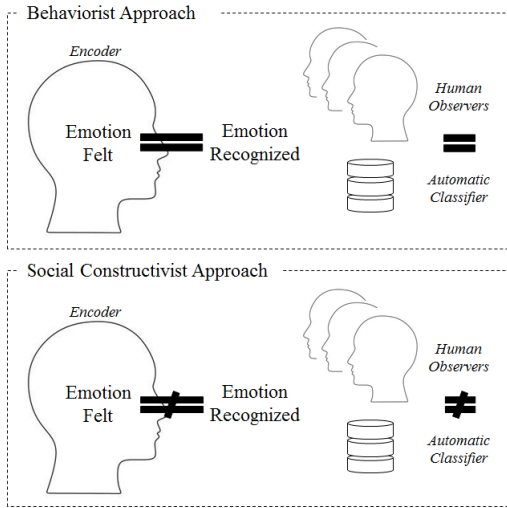


Fig. 1. Comparison of behaviorist and social constructivist approaches according to the expected differences in emotion recognition.

The aim of the current paper is to investigate the coherence between the subjective feeling of emotions and its recognition from facial expressions in order to give credit either to the behaviorist plea or to the constructivist's one. Contrary to most studies using posed and static EFE, the present study focuses on natural EFE. Spontaneous and dynamic facial reactions to emotional elicitations are under consideration to ensure the generalizability of the results to emotional behaviors in ordinary life.

II. METHOD

To evaluate the coherence between subjective feeling of emotions and their recognition from facial expressions, encoders were first recruited to perform an emotion elicitation task while their facial expression was video recorded. Then, the videos of the encoders' faces were shown to human observers and were also analysed by an automatic classifier in order to identify which emotion was displayed.

A. Emotion Elicitation

For the emotion elicitation experiment, 358 French participants (182 females, 176 males, $M_{age} = 47.9$, $SD_{age} = 9.2$) were recruited to perform one out of 11 emotion elicitation tasks designed to trigger a positive, a specific negative or a neutral emotional state. Encoders' face were recorded using an hidden camera resulting 358 front facing 768x576 videos varying from 1s to 1479s (Figure 2). These recordings form the DynEemo database [19].



Fig. 2. Example of a front facing recording synced with the full view of the participant and the elicitation task. This picture is taken from a pilot with projects collaborators and all gave consent for the publication of their photos and videos.

After the emotion elicitation task the encoders rated their subjective feeling on Likert scales from 0 ("not at all") to 5 ("strongly") related to six "basic" emotion labels (*i.e.*, *anger*, *disgust*, *fear*, *happiness*, *surprise* and *sadness*) as well as six "non-basic" emotion labels (*i.e.*, *pride*, *curiosity*, *boredom*, *shame*, *humiliation*, and *disappointment*).

Finally, a debriefing session was performed to ensure that encoders were not durably affected by the emotion elicitation task. The debriefing was also used to check that encoders did not guess the real purpose of the experiment (*e.g.*, being filmed while they were performing an emotional elicitation task) to guarantee facial expressions' genuineness. All encoders gave their agreement on their data and video to be processed for research purposes only.

B. Human Facial Expression Recognition

For the human facial expression recognition method, 1383 student participants were recruited to annotate 232 out of the 358 videos, therefore only the 232 annotated videos will be analysed in this paper. Because videos have different durations, participants had to annotate a series of video corresponding to

30min long in total. Each video was annotated 29 times on average ($SD = 12$).

The annotation of facial expressions was performed on-site using *Oudjat*, a software for designing video annotation experiments [20]. For each video, the annotation procedure followed two steps. First, the participants had to identify the emotional sequences by pressing the space bar of their keyboard to indicate the beginning and the end of the emotional sequences while watching the video. Second, the participants watched each emotional sequence previously identified and labeled the sequence using one of the 12 emotions proposed including six “basic” emotion labels (*i.e.*, *anger*, *disgust*, *fear*, *happiness*, *surprise* and *sadness*) and six “non-basic” emotion labels (*i.e.*, *pride*, *curiosity*, *boredom*, *shame*, *humiliation*, and *disappointment*). They also had the possibility to indicate that the sequence was expressing none of the proposed emotion labels.

This annotation procedure results in a uni-dimensional time-series for each video per human observer identifying for each second of the video which emotion was recognized. Then, time-series corresponding to the same video were aggregated to calculate the proportion of human observers $x_{video_i.label_j.t_k}$ for each second of the video per emotional label (EQ1).

$$x_{video_i.label_j.t_k} = \frac{n_{video_i.label_j.t_k}}{n_{video_k}} \quad (1)$$

where i is one of the 232 videos, j is one of the six “basic” emotion labels, k for each second of the video.

C. Automatic Facial Expression Recognition

The 232 annotated video were processed with Affdex (SDK v3.4.1). Affdex is an automatic facial expression recognition classifier developed and distributed by Affectiva is a spin-off company resulting from the research activities of MIT media lab created in 2009 [21]. Affdex’s algorithm uses Histogram of Oriented Gradient (HOG) features and Support Vector Machine (SVM) classifiers in order to recognize facial expressions. For each video frame, Affdex identifies the probability $p_{video_i.label_j.t_k}$ from 0 to 100 (rescaled to 0 to 1 for the analysis) of the face as expressing one of the six “basic” emotion labels (*i.e.*, *anger*, *disgust*, *fear*, *happiness*, *surprise* and *sadness*) as well as additional psychological states such as *valence*, *engagement* or *contempt*, and facial features such as *cheek raise*, *eye widen* or *jaw drop*.

For both human and automatic recognition, to determine which of the six “basic” emotions can be used to identify each video, the recognition probability for each label by frame was converted into odd ratio by label [22]. The highest sum of each odd ratio time-series defines the label recognized by the automatic classifier (EQ2 for human recognition and EQ3 for automatic recognition).

$$video_i.label = \max \left(\frac{\sum_{k=1}^n x_{video_i.label_j.t_k}}{\sum_{k=1}^n x_{video_i.t_k}} \right) \quad (2)$$

$$video_i.label = \max \left(\frac{\sum_{k=1}^f p_{video_i.label_j.t_k}}{\sum_{k=1}^f p_{video_i.t_k}} \right) \quad (3)$$

where i is one of the 232 videos, j is one of the six “basic” emotion labels, k for each second of the n second video or for each frame of the f frame video.

III. RESULTS

Since encoders’ self-reports, human annotations and the automatic recognition include data on “non-basic” emotion labels and features, the analysis is performed using only the six “basic” emotion labels in order to compare them. The maximum score for self-reports, human annotations and automatic recognition is used to label the video. In case of more than one label obtaining the maximum value, the video is labeled as undetermined.

A. Human Observers’ Accuracy

The overall correlation of recognition and non-recognition between self-reported emotions and human observers recognition is significant but low ($r = .24$, 95% CI [.19, .29], $t(1384) = 9.21$, $p < .001$). In order to identify differences according to the emotional labels, encoders’ subjective feelings are compared with human observers’ recognition in a confusion matrix (Figure 3).

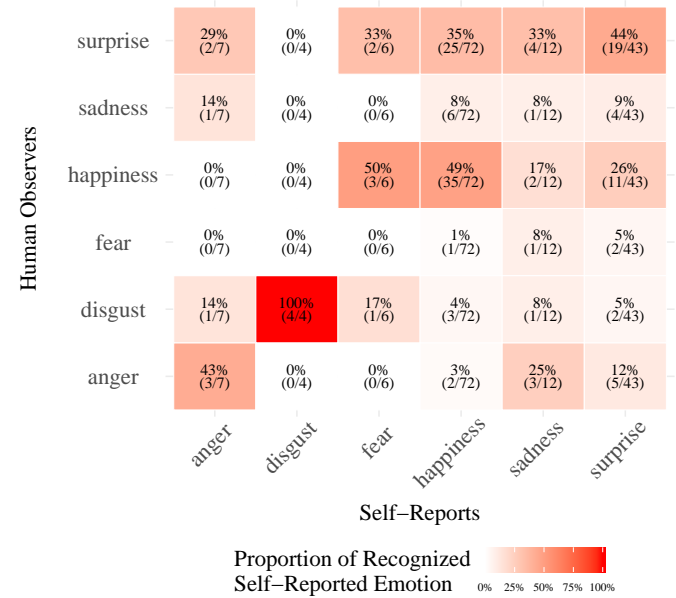


Fig. 3. Confusion matrix between the emotion self-reported as being characteristic of the elicitation with the emotion recognized by the human observers.

Each emotion label used to describe encoders’ self-reported subjective feeling (*i.e.*, the label rated with the highest value) is compared with the emotion labels which were rated with the highest score by human observers. Results of the confusion matrix show a moderate agreement between the emotion felt by the encoder during the elicitation and the emotion recognized by the human observers (Accuracy = 0.43, 95% CI [0.35, 0.52]; Kappa = 0.19) except for *disgust* (100% of the videos self-reported). These results are far from those

classically obtained in the literature for emotional facial expression recognition which ranges between 60% and 80% accuracy. However these results are mostly obtained with static (*i.e.*, pictures) and posed (*i.e.*, displayed by actors) facial expressions using only 6 emotional labels in a forced-choice paradigm.

Interestingly human observers seem to recognize *surprise* expressed in videos where *anger*, *fear* *happiness* and *sadness* was the highest self-reported emotion (respectively 28.6%, 33.3%, 34.7% and 33.3% of the videos self-reported), and in a lower instance *happiness* was recognized in videos where *fear* and *surprise* was the highest self-reported emotion (respectively 50.0% and 25.6% of the videos self-reported).

Sensitivity, specificity, precision and F1 scores for each emotion reveals that *happiness* has the highest coherence ratio whereas *sadness* has the lowest coherence ratio between true positives and false positives (Table I).

TABLE I
HUMAN RECOGNITION ACCURACY METRICS FOR EACH EMOTION.

Emotion	Sensitivity	Specificity	Precision	F1
anger	0.43	0.93	0.23	0.3
disgust	1.00	0.94	0.33	0.5
fear	0.00	0.97	0.00	<i>na.</i>
happiness	0.49	0.78	0.69	0.57
sadness	0.08	0.92	0.08	0.08
surprise	0.44	0.67	0.37	0.4

Note. *na.* values are produced when not enough data are available to compute accuracy indicators.

Accuracy metrics by emotional labels indicate a discrepancy in the ratio of true/false positives. Whereas *happiness* and *disgust* obtain the highest scores, *anger*, *surprise* and *sadness* have the lowest recognition ratio. The underlying effect of expression intensity may explain why *happiness* and *disgust* are easily recognized. *Anger* and *sadness* as non-socially desirable emotions may be have been felt but not expressed.

However, self-reports show a significant proportion of undetermined emotional states (35.2% of the 358 videos) which reveals the potential limit of using 6-points likert scales to measure emotional self-reports. Indeed, encoders can easily score to the maximum for more than one emotion.

B. Automatic Classifier's Accuracy

Similarly to the previous analysis, the overall correlation of recognition and non-recognition revealed a significant but very low coherence between self-reported emotions and automatic classifier's recognition ($r = .12$, 95% CI [.07, .17], $t(1384) = 4.50$, $p < .001$). A confusion matrix was used to compare encoders' subjective feeling with the emotion label recognized by the automatic classifier (Figure 4).

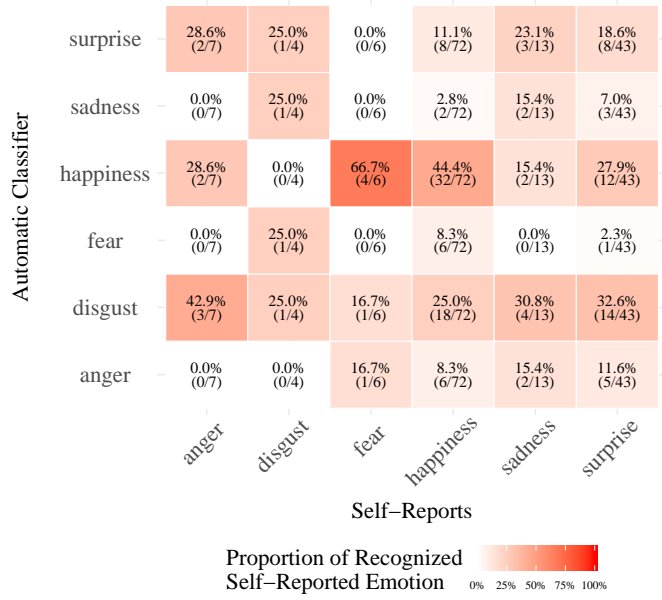


Fig. 4. Confusion matrix of between the emotion self-reported as being characteristic of the elicitation with the emotion recognized by the automatic classifier.

Results obtained for the comparison between emotions self-reported and recognized by the automatic classifier are somewhat similar to the ones with human observers (Table II). Overall, a low agreement between emotion self-reported and emotion recognized by the automatic classifier (Accuracy = 0.3, 95% CI [0.22, 0.38]; Kappa = 0.07) except for *happiness* (44.4% of the video self-reported) is evident.

Surprisingly the automatic classifier incorrectly recognized as *disgust* an significant proportion of videos in which *anger*, *happiness* and *surprise* was the highest self-reported emotion (respectively 42.9%, 25.0% and 32.6% of the videos self-reported). In parallel, the automatic classifier recognized as *happiness* videos in which *fear* and *surprise* was the highest self-reported emotion (respectively 66.7% and 32.6% of the videos self-reported).

TABLE II
AUTONATIC RECOGNITION ACCURACY METRICS FOR EACH EMOTION.

Emotion	Sensitivity	Specificity	Precision	F1
anger	0.00	0.90	0.00	<i>na.</i>
disgust	0.25	0.72	0.02	0.04
fear	0.00	0.94	0.00	<i>na.</i>
happiness	0.44	0.73	0.62	0.52
sadness	0.15	0.95	0.25	0.19
surprise	0.19	0.86	0.36	0.25

Note. *na.* values are produced when not enough data are available to compute accuracy indicators.

A comparable explanation involving the amount of undetermined video based on self-reports can be provided, as

the level of undetermined emotions are very high for the self reports.

C. Comparison Between Human and Automatic Recognition

As previously mentioned, human observers appear to be more accurate than the automatic classifier to recognize an individual's subjective feeling (human observers Accuracy = 0.43; automatic classifier Accuracy = 0.3; $r = .22$, 95% CI [.17, .27], $t(1384) = 8.31$, $p < .001$). However, both make mistakes.

A third confusion matrix is used to compare similarities (diagonal) and differences between human observers and automatic classifier in classifying the six emotion labels (Figure 5).

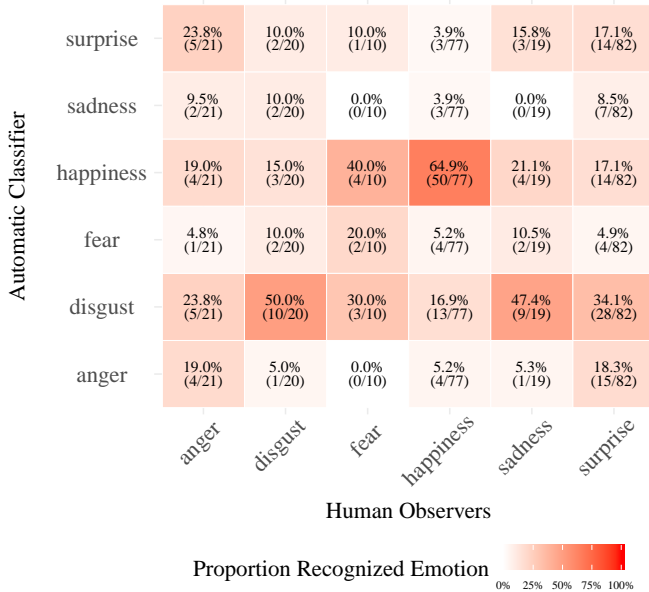


Fig. 5. Proportion of emotion labels classified by human observers which are recognized by the automatic classifier.

The overall agreement between human observers and the automatic classifier is in fact very low (Kappa = 0.18). Except for *happiness* and *disgust* (respectively 64.9% and 50.0% of common labelling), there is no clear common pattern. Moreover, the automatic classifier has a tendency to label as *disgust* videos labeled as *sadness* by human observers, and as *happiness* videos labeled as *fear* by human observers.

IV. DISCUSSION AND CONCLUSION

Despite being one of the most investigated questions in affective science, the coherence between emotion felt and facially displayed on the one hand, and facial expression recognized on the other is a hot topic. To date no clear evidence has been found to definitively solve the questions raised. Yet, with the growing interest of industries and government in monitoring individual's psychological states, this issue is under intense scrutiny. The present research

aims to provide some empirical data to answer some of the questions posed. The faces encoders displayed when confronted with an emotional eliciting task were submitted both to human and to automatic recognition. The criterion for recognition accuracy was the subjective feeling self-reported by the encoder once the elicitation task was carried out. Results first reveal a low coherence between emotion felt and facial expression displayed. Secondly, results show low accuracy rates for both humans and the automatic classifier in identifying the inner emotional states of these encoders based on their facial expressions. Thirdly, human observers prove to be better at recognizing the emotion facially expressed than the automatic recognition tool is.

Such results support the hypothesis advanced by some authors of a low emotion-expression coherence [23]. In many instances, facial displays are not associated with a concordant emotional state, even any emotional state at all [24], [25]. More and more evidence is showing that facial expressions are in reality not expressing emotions [26]. As well as other nonverbal behaviors, facial movements are not only assumed to be determined by emotion but also by various other causes, such as psychological states (e.g., motivations or pain), to say nothing of social context and sociocultural norms [7]. This multiple determination excludes any possibility of drawing an inference from facial activity on the underlying psychological state (emotional or other).

Beyond the present observations showing a weak coherence between subjective feelings and spontaneous facial expressions, this study sheds some light on the controversy between the behaviorist and the constructivist approaches. The first of those two lines of thinking [3] assumes that expressions of emotion are brief and coherent patterns of facial muscle movements that co-vary with discrete subjective experiences. Emotions and their related prototypical facial behavior are universal because they are considered as innate mechanisms allowing individuals to respond adaptively to evolutionary significant events (threats, opportunities...) encountered in the environment. Instead of viewing emotions as natural kinds [12], the constructivist approach supposes that emotions are social constructions. The specific emotion categories identified on the face of another person are the result of categorization. In other words, the emotions that are recognized by the observer are constructed in her/his mind, primarily based on her/his conceptual knowledge of emotion. Therefore, facial movements do not express specific emotions. It is the observer that infers the emotional meaning of the facial expression. As a consequence, one can predict from the first line of thinking that individuals' emotional subjective feeling should be correlated to the recognition of facial expressions from both human observers and automatic classifiers whereas if emotions are social constructs, as stated by the second line of thinking, human observer's should be better at perceiving emotions expressed on the face than automatic classifiers.

Present results plead in favor of the latter stance. They show that human observers are more accurate than the automatic

recognition tool to identify an individual's subjective feeling on the basis of their face. Moreover, mistakes made by human observers look less arbitrary to the ones made by the classifier. For instance, even if a mix-up between disgust and anger is sometimes reported in recognition studies, confusions such as the ones produced by the classifier have never been noted for human observers. A possible explanation is that human observers are assessing dynamic expressive sequences. They have access to the full video which provides a kinetic context whereas the automatic classifier only assesses the videos frame by frame. Further research is needed to shed light on this issue.

Some limitations should be stated, notably regarding the use of self-reports to evaluate encoders' subjective feelings. Accessing the inner subjective feeling can be biased if not impossible. Moreover the procedure used for human observation can also be open to dispute. Instead of asking the human annotators to provide a unique label, a more subtle approach was chosen to mimic results provided by the automatic classifier. Whereas this paradigm is longer and more complicated, it can lead to more robust results in reducing the forced-choice bias [27]. However this procedure can also reduce the human observers' accuracy. In this regard, the results of the human observation could have been more ambiguous because it is not the natural way that people are inferring meaning from facial expressions. An alternative explanation relies in reducing the recognition bias involved in the classic recognition paradigm. Classic forced-choice paradigms obtain artificially high results, thus by using a more evolved approach observers' accuracy may have been lowered.

Considering the above, the results provide additional evidence that an individual's subjective feeling can not be inferred from facial expressions and in our case invalidate the hypothesis of hardwired emotions unambiguously displayed on the face. Even if emotions were hardwired, in everyday life one does not observe prototypical facial expressions and therefore research should be focused on analysing non-prototypical facial expressions. Advancements in identifying "non-basic" emotion labels [28] as well as non-prototypical facial expression have been made in the development of automatic facial expression recognition tools. However, these results suggest that automatic facial expression recognition tools should merely evaluate facial morphology features such as action units (already evaluated in OpenFace [29], Affectiva's Affdex [21] or Vicar Vision's FaceReader [30] to name a few) rather than inferring supposedly emotional or affective states.

V. ACKNOWLEDGMENT

The authors would like to thank Brigitte Meillon and Jean Michel Adam who developed the software used to collect and preprocess human observer results.

VI. REFERENCES

- [1] D. Dupré, N. Andelic, G. Morrison, and G. McKeown, "Accuracy of three commercial automatic emotion recognition systems across different individuals and their facial expressions," in *Proceedings of the international conference on pervasive computing and communications*, 2018, pp. 627–632.
- [2] C. Darwin, *The expression of the emotions in man and animals*. London, UK: John Murray, 1872.
- [3] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, nos. 3–4, pp. 169–200, 1992.
- [4] F. de Waal, *Mama's last hug: Animal emotions and what they teach us about ourselves*. London, UK: Granta Books, 2019.
- [5] P. Ekman, "The directed facial action task," in *Handbook of emotion elicitation and assessment*, J. A. Coan and J. J. B. Allen, Eds. New York, NY: Oxford University Press, 2007, pp. 47–53.
- [6] P. Ekman and K. G. Heider, "The universality of a contempt expression: A replication," *Motivation and Emotion*, vol. 12, no. 3, pp. 303–308, 1988.
- [7] P. Ekman et al., "Universals and cultural differences in the judgments of facial expressions of emotion," *Journal of Personality and Social Psychology*, vol. 53, no. 4, pp. 712–717, 1987.
- [8] R. E. Kraut and R. E. Johnston, "Social and emotional messages of smiling: An ethological approach," *Journal of Personality and Social Psychology*, vol. 37, no. 9, pp. 1539–1553, 1979.
- [9] J. I. Durán, R. Reisenzein, and J.-M. Fernández-Dols, "Coherence between emotions and facial expressions," pp. 107–129, 2017.
- [10] J. R. Averill, "A constructivist view of emotion," in *Theories of emotion*, R. Plutchik and H. Kellerman, Eds. New York, NY: Academic Press, 1980, pp. 305–339.
- [11] L. F. Barrett, *How emotions are made: The secret life of the brain*. Boston, MA: Houghton Mifflin Harcourt, 2017.
- [12] L. F. Barrett, "The theory of constructed emotion: An active inference account of interoception and categorization," *Social Cognitive and Affective Neuroscience*, vol. 12, no. 1, pp. 1–23, 2017.
- [13] C. Crivelli and A. J. Fridlund, "Facial displays are tools for social influence," *Trends in Cognitive Sciences*, vol. 22, no. 5, pp. 388–399, 2018.
- [14] N. H. Frijda and A. Tcherkassof, "Facial expressions as modes of action readiness," in *The psychology of facial expression*, J. A. Russell and J.-M. Fernández-Dols, Eds. Cambridge, UK: Cambridge University Press, 1997, pp. 78–102.
- [15] A. J. Fridlund, *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press, 1994.
- [16] K. A. Lindquist and M. Gendron, "What's in a word? Language constructs emotion perception," *Emotion Review*, vol. 5, no. 1, pp. 66–71, 2013.
- [17] P. M. Niedenthal, A. Wood, M. Rychlowska, and S. Korb, "Embodied simulation in decoding facial expression," in *The science of facial expression*, J.-M. Fernández-Dols and J. A. Russell, Eds. New York, NY: Oxford University Press, 2017, pp. 397–414.
- [18] C. M. Doyle and K. A. Lindquist, "Language and emotion: Hypotheses on the constructed nature of emotion perception," in *The science of facial expression*, J.-M. Fernández-Dols and J. A. Russell, Eds. New York, NY: Oxford University Press, 2017, pp. 415–432.
- [19] A. Tcherkassof, D. Dupré, B. Meillon, N. Mandran, M. Dubois, and J.-M. Adam, "DynEmo: A video database of natural facial expressions of emotions," *The International Journal of Multimedia & Its Applications*, vol. 5, no. 5, pp. 61–80, 2013.
- [20] D. Dupré et al., "Oudjat: A configurable and usable annotation tool for the study of facial expressions of emotion," *International Journal of Human-Computer Studies*, vol. 83, pp. 51–61, 2015.
- [21] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. el Kaliouby, "AFFDEX sdk: A cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of the chi conference on human factors in computing systems*, 2016, pp. 3723–3726.
- [22] P. Dente, D. Küster, L. Skora, and E. Krumhuber, "Measures and metrics for automatic emotion classification via facet," in *Proceedings of the conference on the study of artificial intelligence and simulation of behaviour*, 2017, pp. 160–163.
- [23] A. Kappas, "What facial activity can and cannot tell us about emotions," in *The human face*, M. Katsikitis, Ed. Boston, MA: Springer, 2003, pp. 215–234.
- [24] G. Bonanno and D. Keltner, "The coherence of emotion systems: Comparing 'on-line' measures of appraisal and facial expressions, and self-report," *Cognition and Emotion*, vol. 18, no. 3, pp. 431–444, 2004.
- [25] J.-M. Fernández-Dols and C. Crivelli, "Emotion and expression: Naturalistic studies," *Emotion Review*, vol. 5, no. 1, pp. 24–29, 2013.
- [26] G. J. McKeown, "The analogical peacock hypothesis: The sexual selection of mind-reading and relational cognition in human communication," *Review of General Psychology*, vol. 17, no. 3, pp. 267–287, 2013.
- [27] J. A. Russell, "Forced-choice response format in the study of facial expression," *Motivation and Emotion*, vol. 17, no. 1, pp. 41–51, 1993.

[28] D. McDuff, “Discovering facial expressions for states of amused, persuaded, informed, sentimental and inspired,” in *Proceedings of the international conference on multimodal interaction*, 2016, pp. 71–75.

[29] T. Baltrušaitis, P. Robinson, and L.-P. Morency, “Openface: An open source facial behavior analysis toolkit,” in *Proceedings of the winter*

conference on applications of computer vision, 2016, pp. 1–10.

[30] M. Den Uyl and H. Van Kuilenburg, “The facereader: Online facial expression recognition,” in *Proceedings of measuring behavior*, 2005, vol. 30, pp. 589–590.