Are facial expressions the genuine display of individuals’ subjective feeling? A comparison of human and automatic recognition.

While it has been taken for granted in the development of several automatic facial expression recognition tools, the question of the link 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 are socially constructed and facial expression as social messages that are not related to emotions. In order to evaluate the link between the subjective feeling of emotions and their recognition based on facial expression, 232 videos of participants recruited to perform an emotion elicitation task were annotated by 1383 human observers as well as by an automatic facial expression classifier. The results show a low accuracy of human observers and of the automatic classifier to infer from the facial expression the subjective feeling of the participants recorded. Based on these results, the hypothesis of genetically hardwired emotion genuinely display is difficult to support whereas the idea of emotion socially constructed and facial expression as display of social messages appears to be more likely. Then, the way to infer emotional and mental state based on facial expressions should be questioned.

# Introduction

With the development of commercial automatic facial expression recognition tools (see Dupré et al. 2018 for a non-exhaustive list of available tools), industries and governments are gradually implementing this technology in order to track humans’ emotions in various scenarios (e.g., marketing, healthcare, automotive to name a few). This technoloogy rests on the premise that facial expressions provide a direct access to individuals’ subjective feeling. Even if this premise is central to the modern mainstream approach of human emotion, recent research in affective science are challenging it. After presenting the arguments supporting both sides, an experiment testing these hypotheses will be presented and its results analyzed in order to provide empirical evidence to contribute to answer the question.

## The link between emotion and facial expression through the behaviorist approach

Based on the behaviorist approach initiated by Darwin in *The Expression of the Emotions in Man and Animals* (Darwin 1872), facial expression are conceived as a genuine display of individuals inner emotional state. This hypothesis is used as a basis for the Basic Emotion Theory (BET) which states that a set of six emotions are universally displayed and are genetically hardwired not only in humans (Ekman 1992) but also in different animal species (Waal 2019). According to this view, “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.” (Ekman 2007, p49). To cope with critics, several amendments have been made to the BET, increasing the number of basic emotions from six to seven (Ekman and Heider 1988) as well as adding the concept of “display rules” to explain cultural differences in the management of facial expressions (Ekman et al. 1987). Even if this theory obtained a popular support, it fails to explain how individuals can feel emotions without expressing them and how individuals can express emotions without feeling them, cases for which display rules cannot be applied to (Kraut and Johnston 1979, @duran2017coherence).

## The link between emotion and facial expression through the social constructivist approach

Detractors of the Basic Emotion Theory are perceiving emotion not as genetically hardwired but as a learnt association between a given situation and an appropriate response (Averill 1980, @barrett2017emotions). For the tenants 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” (Barrett 2017b, p12). Following this assumption, faces are used as tools to display signals in social interactions (Crivelli and Fridlund 2018). These signals can convey individuals’ motivations and readiness (Frijda and Tcherkassof 1997) or social messages (Fridlund and Rosenberg 1995). Therefore, facial expressions are thought as behaviors which meaning is inferred by the observer.

This current paper investigates the link between the subjective feeling of emotions and their recognition from facial expressions. If emotions are hardwired, 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, no correlation between subjective feeling and facial expression recognition should be observed.

# Method

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

## 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, negative or neutral emotional state. Participants’ face were recorded using an hidden camera resulting 358 front facing 768x576 videos varying from 1s to 1479s (Figure 1). These recordings are constituting the DynEmo database (see Tcherkassof et al. 2013 for a full description of tasks and procedure).



Figure 1: 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 a consent for the publication of their photos and videos.

After their emotion elicitation task the participants had to rate their subjective feeling during the task on likert scales from 0 (“not at all”) to 5 (“strongly”) according 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 perform to ensure that participants were not durably affected by the emotion elicitation task. The debriefing was also used to check that participants 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 the participants gave their agreement on their data and video to be processed for research purpose only.

## Human Facial Expression Recognition

For the human facial expression recognition method, 1383 student participants were recruited to annotate 232 out of the 358 video, therefore only the 232 annotated videos will be analysed in this paper. Each participants had to annotate between 1 and 177 videos resulting that 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 (Dupré et al. 2015). For each video, the annotation procedure hat 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 had to label 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 not expressing one of the proposed emotion.

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 for each second of the video per emotional label. The sum of each label proportion per second was used as a score to determine which labels corresponds to the overall video (i.e., the highest score). In case of more than one label having the maximum value, the emotion is described as undetermined.

## 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 (McDuff et al. 2016). 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 identify the probability 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*.

To determine which of the six “basic” emotion can be used to identify each video, the recognition probability for each label by frame was converted into odd ratio by frame (Dente et al. 2017). The highest sum of each odd ratio time-series defines the label recognized by the automatic classifier.

# Results

Whereas the self-reports, the 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.

## Correlation between self-report and human facial expression recognition

Participants’ subjective feeling is compared with human observers recognition in a confusion matrix (Figure 2). Each emotion label used to describe participants subjective feeling (i.e., the label rated with the highest value) are compared with the emotion labels which was rated with the highest score by human observers

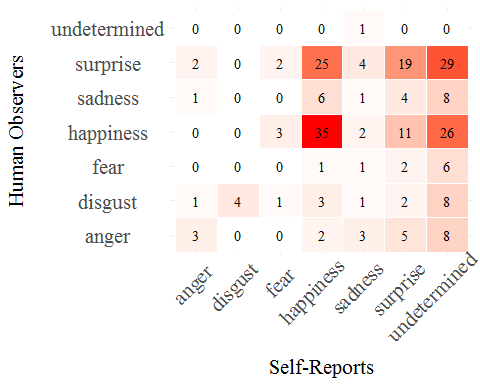


Figure 2: Confusion matrix of between the emotion self-reported as being characteristic of the elicitation with the emotion recognized by the human observers.

The result of the confusion matrix show a low agreement between emotion felt during the elicitation and emotion recognized by the human annotators (Accuracy = 0.27, 95%CI[0.21,0.33]; Kappa = 0.11) except for *happiness* (48.6% of the video self-reported), *surprise* (44.2% of the videos self-reported) and *disgust* (100% of the videos self-reported). Sensitivity, specificity, precision and F1 score for each emotion can be found Table 1. Interestingly human annotators seem to recognize as *surprise* videos in which *happiness* was the highest self-reported emotion (34.7% of the videos self-reported), and in a lower instance *happiness* videos in which *surprise* was the highest self-reported emotion (25.6% of the videos self-reported).

(#tab:confusionTable\_sr\_hr)

*Table 1: Agreement accuracy metrics for each emotion.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Emotion | Sensitivity | Specificity | Precision | F1 |
| anger | 0.43 | 0.92 | 0.14 | 0.21 |
| disgust | 1.00 | 0.93 | 0.20 | 0.33 |
| fear | 0.00 | 0.96 | 0.00 |  |
| happiness | 0.49 | 0.73 | 0.45 | 0.47 |
| sadness | 0.08 | 0.91 | 0.05 | 0.06 |
| surprise | 0.44 | 0.67 | 0.23 | 0.31 |
| undetermined | 0.00 | 0.99 | 0.00 |  |

However, the self-report show a very high proportion of undetermined emotional states which reveals the potential limit of using 6-points likert scales for which the participants can easily score to the maximum for more than one emotion.

## Correlation between self-report and automatic facial expression recognition

Similarly to the previous analysis, a confusion matrix was used to compare participants subjective feeling with the emotion label recognized by the automatic classifier (Figure 3).

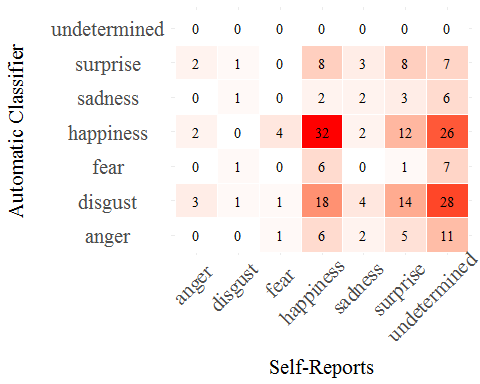


Figure 3: Confusion matrix of between the emotion self-reported as being characteristic of the elicitation with the emotion recognized by the automatic classifier.

Results obtained for the comparison between emotions self-reported and recognized by the automatic classifier are similar to the ones with human observers (Table 2). Overall there is a low agreement between emotion self-reported and emotion recognized by the automatic classifier (Accuracy = 0.19, 95%CI[0.14,0.24]; Kappa = 0.05) except for *happiness* (44.4% of the video self-reported) and *surprise* (18.6% of the videos self-reported). Surprisingly the automatic classifier incorrectly recognized as *disgust* an important proportion of videos in which *happiness* was the highest self-reported emotion (25.0% of the videos self-reported). In parallel, the automatic classifier recognized as *happiness* and *disgust* videos in which *surprise* was the highest self-reported emotion (respectively 27.9% and 32.6% of the videos self-reported).

(#tab:confusionTable\_sr\_ar)

*Table 2: Agreement accuracy metrics for each emotion.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Emotion | Sensitivity | Specificity | Precision | F1 |
| anger | 0.00 | 0.89 | 0 |  |
| disgust | 0.25 | 0.70 | 0.01 | 0.03 |
| fear | 0.00 | 0.93 | 0 |  |
| happiness | 0.44 | 0.71 | 0.41 | 0.43 |
| sadness | 0.15 | 0.94 | 0.14 | 0.15 |
| surprise | 0.19 | 0.89 | 0.28 | 0.22 |
| undetermined | 0.00 | 1.00 |  |  |

A comparable explanation can be provided as the level of undetermined emotion are very high for the self reports.

## Comparison between human and automatic recognition

As previously mentioned, the accuracy of humans observers and the automatic classifier have some similarities. In order to compare which recognition has the highest accuracy a Receiver Operating Characteristic (ROC) curve was calculated (Figure 4).

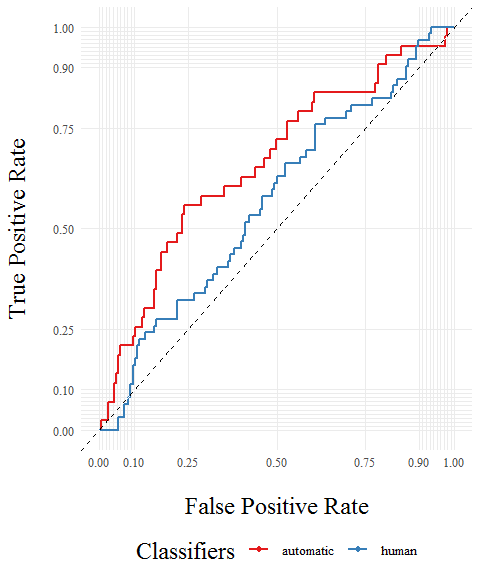


Figure 4: ROC curve comparing the accuracy in inferring subjective feelings from facial expressions by human observers and automatic recognition.

The ROC curve and its Area Under the Curve (AUC) values shows that the automatic classifier are more accurate than human observers in inferring subjective feeling from facial expressions (human AUC = 0.57 *vs.* automatic AUC = 0.66).

# Conclusion

Despite being one on the most investigated question in affective science, the link between emotion felt and facial expression is a hot topic and no clear evidence have been found to definitely answer it. However, with the growing interest of industries and government to monitor individual’s psychological states, evidences are showing that facial expressions are in reality not expressing emotions (McKeown 2013). This research aimed to provide some empirical data to the question. The subjective feeling of participants was compared with human recognition on one side and automatic recognition on the other side. The results reveals a low accuracy for both humans and automatic classifier to accurately identify the inner emotional states of these individuals based on their facial expressions. Moreover it appeared that automatic classifier obtain a higher accuracy to recognize the corresponding emotion label than humans observers are. A possible explaination is that human oberservers are usually not using emotion label to recognize emotion and this type of paradigm provides non accurate results.

Some limitations to this process should be stated over the use of self-reports to evaluate individual’s subjective feelings. Accessing the inner subjective feeling can be biased if not impossible. Moreover the procedure used for human annotation can also be open to dispute. Instead of asking the human annotators to provide an 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 biais (Russell 1993). However this procedure can also reduce the human obersvers accuracy. In this regard, the results of the human annotation could have been more ambiguous because it is not the natural way that people are inferring meaning from facial expressions. An alternative explaination relies in reducing the recognition biais involved in the classic recognition paradigm. Classic forced-choice paradigms obtain artificial high results, thus by using a more evolved approach observers’ accuracy may have been reduced.

Considering the above, the results provides an additional evidence that individuals’ subjective feeling can not be inferred from facial expressions and in our case invalidate the hypothesis of hardwired emotions. 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 (McDuff 2016) as well as non-prototypical facial expression have been made in the developpement of automatic facial expression recognition tools. However this result suggests that automatic facial expression recognition tools should evaluate facial morphology features such as action units (e.g., OpenFace Baltrušaitis, Robinson, and Morency 2016) rather than inferring potential emotional or affective states.

# Acknowledgment

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

# References

Averill, James R. 1980. “A Constructivist View of Emotion.” In *Theories of Emotion*, 305–39. Elsevier.

Baltrušaitis, Tadas, Peter Robinson, and Louis-Philippe Morency. 2016. “Openface: An Open Source Facial Behavior Analysis Toolkit.” In *2016 Ieee Winter Conference on Applications of Computer Vision (Wacv)*, 1–10. IEEE.

Barrett, Lisa Feldman. 2017a. *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt.

———. 2017b. “The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization.” *Social Cognitive and Affective Neuroscience* 12 (1): 1–23.

Crivelli, Carlos, and Alan J Fridlund. 2018. “Facial Displays Are Tools for Social Influence.” *Trends in Cognitive Sciences* 22 (5): 388–99.

Darwin, Charles. 1872. *The Expression of the Emotions in Man and Animals*. London, UK: John Murray, London UK.

Dente, Pasquale, Dennis Küster, Lina Skora, and E Krumhuber. 2017. “Measures and Metrics for Automatic Emotion Classification via Facet.” In *Proceedings of the Conference on the Study of Artificial Intelligence and Simulation of Behaviour (Aisb)*, 160–63.

Dupré, Damien, Daniel Akpan, Elena Elias, Jean-Michel Adam, Brigitte Meillon, Nicolas Bonnefond, Michel Dubois, and Anna Tcherkassof. 2015. “Oudjat: A Configurable and Usable Annotation Tool for the Study of Facial Expressions of Emotion.” *International Journal of Human-Computer Studies* 83: 51–61.

Dupré, Damien, Nicole Andelic, Gawain Morrison, and Gary McKeown. 2018. “Accuracy of Three Commercial Automatic Emotion Recognition Systems Across Different Individuals and Their Facial Expressions.” In *2018 Ieee International Conference on Pervasive Computing and Communications Workshops (Percom Workshops)*, 627–32. IEEE.

Durán, Juan I, Rainer Reisenzein, and José-Miguel Fernández-Dols. 2017. “Coherence Between Emotions and Facial Expressions.” *The Science of Facial Expression*, 107–29.

Ekman, Paul. 1992. “An Argument for Basic Emotions.” *Cognition & Emotion* 6 (3-4): 169–200.

———. 2007. “The Directed Facial Action Task.” *Handbook of Emotion Elicitation and Assessment* 47: 53.

Ekman, Paul, Wallace V Friesen, Maureen O’sullivan, Anthony Chan, Irene Diacoyanni-Tarlatzis, Karl Heider, Rainer Krause, et al. 1987. “Universals and Cultural Differences in the Judgments of Facial Expressions of Emotion.” *Journal of Personality and Social Psychology* 53 (4): 712.

Ekman, Paul, and Karl G Heider. 1988. “The Universality of a Contempt Expression: A Replication.” *Motivation and Emotion* 12 (3): 303–8.

Fridlund, Alan J, and Erika L Rosenberg. 1995. “Human Facial Expression: An Evolutionary View.” *Nature* 373 (6515): 569–69.

Frijda, Nico H, and Anna Tcherkassof. 1997. “Facial Expressions as Modes of Action Readiness.” *The Psychology of Facial Expression*, 78–102.

Kraut, Robert E, and Robert E Johnston. 1979. “Social and Emotional Messages of Smiling: An Ethological Approach.” *Journal of Personality and Social Psychology* 37 (9): 1539.

McDuff, Daniel. 2016. “Discovering Facial Expressions for States of Amused, Persuaded, Informed, Sentimental and Inspired.” In *Proceedings of the 18th Acm International Conference on Multimodal Interaction*, 71–75. ACM.

McDuff, Daniel, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby. 2016. “AFFDEX Sdk: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit.” In *Proceedings of the 2016 Chi Conference Extended Abstracts on Human Factors in Computing Systems*, 3723–6. ACM.

McKeown, Gary J. 2013. “The Analogical Peacock Hypothesis: The Sexual Selection of Mind-Reading and Relational Cognition in Human Communication.” *Review of General Psychology* 17 (3): 267–87.

Russell, James A. 1993. “Forced-Choice Response Format in the Study of Facial Expression.” *Motivation and Emotion* 17 (1): 41–51.

Tcherkassof, Anna, Damien Dupré, Brigitte Meillon, Nadine Mandran, Michel Dubois, and Jean-Michel Adam. 2013. “DynEmo: A Video Database of Natural Facial Expressions of Emotions.” *The International Journal of Multimedia & Its Applications* 5 (5): 61–80.

Waal, Frans de. 2019. *Mama’s Last Hug: Animal Emotions and What They Teach Us About Ourselves*. Granta Books.