Bridging the Gap Between Automated and Human Facial Emotion Perception

Derek Stratton
University of Nevada Reno
derekstratton@nevada.unr.edu

Emily Hand
University of Nevada Reno
emhand@unr.edu

Abstract

Understanding the complex relationship between emotions and facial expressions is important for both psychologists and computer scientists. A large body of research in psychology investigates facial expressions, emotions, and how emotions are perceived from facial expressions. As computer scientists look to incorporate this research into automatic emotion perception systems, it is important to understand the nature and limitations of human emotion perception. These principles of emotion science affect the way datasets are created, methods are implemented, and results are interpreted in automated emotion perception. This paper aims to distill and align prior work in automated and human facial emotion perception to facilitate future discussions and research at the intersection of the two disciplines.

1. Introduction

Emotions are a core part of the human experience. From the pleasant joy of engaging in a favorite hobby to the sorrow of struggling through a difficult period in life, emotions add color to the many experiences we all have. All kinds of emotions have been invoked in art, literature, music, speech, dance, and countless other mediums throughout history [73]. Emotions can change our perception of the world and influence the actions we take every day [16]. A large focus of emotion research is the relationship between emotions and facial expressions. Facial expressions are viewed as a key to nonverbal communication and for expressing emotion [2].

Despite the omnipresence of emotion and expressions throughout life, there is a lack of scientific consensus regarding many key questions. A large number of theories and body of research on emotion emerged in the 20th and 21st centuries [37, 69]. The study of facial emotion perception in psychology is continuously developing and dynamic. This complexity can make it challenging for researchers outside the field to understand emotion science with its many nuanced and contradictory ideas.

As psychologists have increased their interest in studying emotion, so too have computer scientists, focusing on developing automated emotion perception systems [50]. Automatic emotion perception has applications in assistive technology, human-computer interaction, as well as many others [49]. Unfortunately, the nuances of human emotion perception are often lost when translated automated emotion perception. Some of this stems from the practical requirements of these automated systems (e.g. limited data) while some stems from a fundamental lack of knowledge of the psychological perspective.

The goal of this work is to align both perspectives (psychology and computer science, i.e. human and automated) of emotion perception, facilitating future discussions and research in this interdisciplinary area. It is essential that facial emotion perception be researched and discussed through an interdisciplinary lens as findings can help both fields of psychology and computer science.

The remainder of our paper is organized as follows. Section 2 details prominent emotion theories, while section 3 details the models used to classify emotions. Section 4 focuses on facial expressions and their relationship to emotion. Section 5 provides a discussion of the psychology and computer science perspectives with suggestions for future research and section 6 summarizes and concludes our work.

2. Emotion Theories

Emotion theory has intrigued philosophers, academics, and psychologists for hundreds of years [37]. Psychologists have contemplated many questions about the fundamental properties of emotion, aiming to determine what functions emotions serve, how emotions are related to memory and what separates emotions from mood and temperament [1, 39]. In this paper, we focus on one question: "Are there basic emotions?" Basic emotion theory states that there is a small set of emotions that are innate to humans. The alternative explanation – emotion construction theory – is that humans have created categories of emotion to help better understand the subject. The underlying emotion theory dictates the choice of emotion model and therefore has significant impacts on downstream research tasks.

2.1. Basic Emotion Theory

Basic Emotion Theory states that there exists a small set of distinct, fundamental emotions that are biologically innate to humans. The concept of a facial expression is important when describing a basic emotion because many supporters of Basic Emotion Theory directly link emotions and facial expressions. Basic emotions are described as having "distinct physiology" and "brief duration" [28]. Basic emotions are usually grouped into a small, discrete set that are thought to be hardwired into the brain [19].

Charles Darwin is often considered the inspiration for the Basic Emotion Theory, having stated that basic emotions are innate and based on specific facial and bodily expressions [23,37]. He asserted that these emotions and emotional expressions came about due to their adaptive functions. For example, fear can facilitate the survival of the organism against a hostile attack. Darwin's perspective later influenced the work of other psychologists [24, 45, 74].

A cross-cultural study of the Fore Tribesmen in Papua New Guinea was one of the first major studies used to provide large-scale evidence for basic emotions [33]. They found that the study subjects were able to match pictures of facial expressions to a set of 6 discrete emotion labels, supporting the idea that basic emotions exist across cultures. Analyses of the many cross-cultural studies on emotion have found both consistency in the expression and recognition of emotion across cultures, and variance that can be explained by cultural and other factors [72].

Some have shown that there is neurological evidence for basic emotions, but there is not a one-to-one mapping between a brain responses and basic emotions [20]. While there is undisputed evidence that there are some biological underpinnings for emotion, there is still much debate over what this means. Those who support Basic Emotion Theory believe that there exists some mapping between neurological responses and basic emotions. Others believe that this neurobiological evidence simply points to similar core patterns that are observed from emotion words [67].

2.2. Emotions as Social Constructs

Social Constructionists argue that emotions do not exist in a discrete set of biologically-innate categories, but instead that societies create emotion categories as a way to better understand affect and facilitate communication about feelings [7]. Emotion and affect are often used interchangeably. Affect is "any experience of feeling or emotion," and "both mood and emotion are considered affective states" [76]. These conceptions of emotion need not be universal, which is an important distinction from Basic Emotion Theory. Constructionists often emphasize a mix of both neurological mechanisms and social factors like context and culture to explain the variability of emotion [6].

William James introduced ideas that would later serve as the foundation for Constructed Emotion Theory, describing emotion as a product of elementary processes [37, 47, 53]. Constructed Emotion Theory gained popularity in the 21st century [26, 37].

[68] distinguishes two components of an emotion: core affect and prototypical emotion episodes . Core affect is an internal state characterized by "elemental processes of pleasure and activation," and prototypical emotion episodes are rare, complex sequences of events that match a typical emotion category like anger or fear. [67] explains that emotion categories are not clearly defined, and only some emotional episodes would fit well with a prototypical category. [5] posits that concepts like fear, anger, and sadness are social constructs used to classify these emotional episodes since people experience similar patterns of emotion [5]. [51] found the combination of core affect and conceptual knowledge to be important in the experience of fear. [4] used neuroimaging studies to show that different neural structures are activated by different emotion-related stimuli, showing evidence for complex processes for constructing emotion.

Many believe that a combination of innate and constructed emotions exist. [44] states that Basic Emotion Theory is still valid if basic emotions are restricted to emotions that are characterized by "evolutionary adaptions that are involuntarily and automatically triggered." Some take research supporting Constructed Emotion Theory as evidence for Basic Emotion Theory, acknowledging that there are more categories of emotion that have more complex relationships with the brain than previously thought [48].

The most important takeaway from emotion theory is that there is support for both innate and environmental explanations of emotion. Researchers in automated emotion perception should be aware of these theories and how they impact critical research design choices including data, models and applications.

3. Emotion Models

A host of models have been created that attempt to provide a system for comparing and measuring emotions [17]. Emotion models can be divided into two broad classification approaches: categorical – representing the space of all emotions as a finite set – and dimensional representing emotions by continuous values on multiple axes [17]. These two classes of emotion models align with Basic Emotion Theory and Constructed Emotion Theory, respectively.

Choosing an emotion model is dependent on the emotion theory underlying the work and the problem the model is applied to. Categorical models consist of semantic categories, which are generally more intuitive. Dimensional models have a larger range of representations and are better at quantifying the relationship between different emotional states.

Table 1. Categorical Emotion Models. Emotions common to all models are shown in bold. It should also be noted that each model contains either ”happiness” or ”joy”.

Author(s)	Count	Emotions
Ekman [27]	6	**Anger**, Disgust, **Fear**, Happiness, **Sadness**, **Surprise**
Ekman and Friesen [31]	7	**Anger**, Contempt, Disgust, **Fear**, Happiness, **Sadness**, **Surprise**
Ekman [29]	15	Amusement, **Anger**, Contempt, Contentment, Disgust, Embarrassment, Excitement, **Fear**, Guilt, Happiness, Pride in achievement, Relief, **Sadness**, Satisfaction, Sensory pleasure, Shame, **Surprise**
Plutchik [60]	8	**Anger**, Anticipation, Disgust, **Fear**, Joy, **Sadness**, **Surprise**, Trust
Parrott [58]	6	**Anger**, **Fear**, Joy, Love, **Sadness**, **Surprise**
Cowen and Kelter [22]	27	Admiration, Adoration, Aesthetic appreciation, Amusement, **Anger**, Anxiety, Awe, Awkwardness, Boredom, Calmness, Confusion, Craving, Disgust, Empathic pain, Entrancement, Excitement, **Fear**, Horror, Interest, Joy, Nostalgia, Relief, Romance, **Sadness**, Satisfaction, Sexual Desire, **Surprise**

3.1. Categorical Emotion Models

Categorical emotion models classify emotions into distinct categories, which aligns well with Basic Emotion Theory. These models started with very few categories, but compound facial expressions have been studied to create more emotion categories [25]. Table 1 details various categorical emotion models.

Ekman's original emotion model contains 6 basic emotions [27]. This model was revised two more times to include a total of 15 basic emtions [29, 31]. Beyond Ekman's models, some categorical models have additional structure. Robert Plutchik describes emotions on a wheel, with opposite emotions being on opposite sides of the wheel [60]. Figure 1 shows Plutchik's emotion wheel, with 8 basic emotions that have mild forms, intense forms, and combinations [59].

Parrott created a tree structure of emotions based on finding the prototypicality of various emotion keywords rated by students [58]. The first 2 layers of the tree are shown in figure 2. The tree-like structure emphasizes how the differences between emotions can be distinct or subtle.

Cowen and Kelter utilized responses to emotion-eliciting videos to created a discrete list of 27 emotions, noting that emotion states have fuzzy boundaries [22]. Their emotion model is visualized in figure 3.

Automated emotion perception has relied heavily on categorical emotion models, and specifically Ekman's original model with 6 emotions. Using other categorical models is an obvious next step in the automated perception of more subtle emotions.

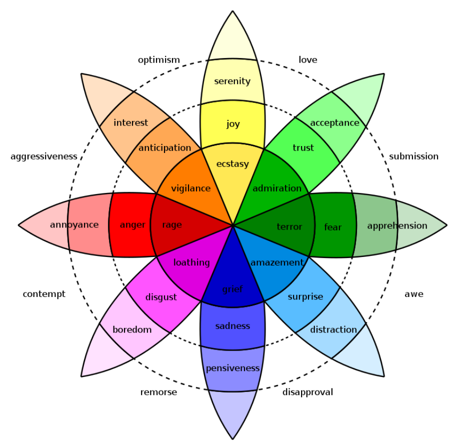

Figure 1. Plutchik's Wheel of Emotions shows 8 basic emotions represented by leaves on a wheel. Words closer to or farther from the center represent higher or lower intensities of the emotion, respectively. Adjacent petals represent similar emotions and opposing petals represent opposing emotions. The words between the petals describe emotions related to the adjacent petals [60]

Table 2. Dimensional Emotion Models

Author(s)	Count	Dimensions
Russell [66]	2	Valence and Arousal
Watson and Tellegen [77]	2	Positive-Affect and Negative-Affect
Mehrabian [56]	3	Pleasure, Arousal, and Dominance

3.2. Dimensional Emotion Models

Dimensional emotion models represent emotions with a set of real values scored on independent axes aligning well with Constructed Emotion Theory. Table 2 details common dimensional models.

James Russell introduced the first dimensional model – the Circumplex Model [66]. The Circumplex Model was constructed by plotting various emotion keywords on a circle that contained two real-valued axes: valence and

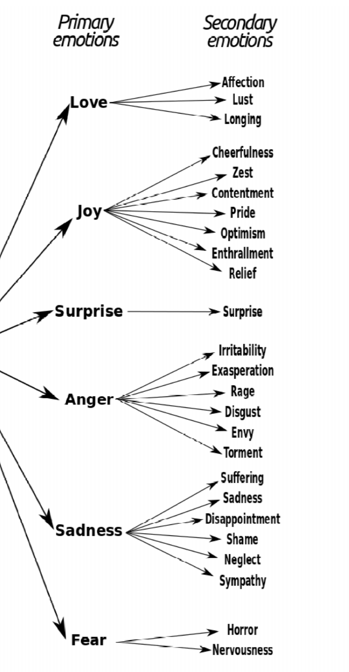

Figure 2. Parrott's Tree of Emotions define a list of 6 primary emotions, and various secondary emotions that stem from a primary emotion. Tertiary emotions are not shown [12].

Figure 3. Cowen and Kelter's Mapping of Emotional Videos plotted with t-SNE. The colors of the points represent the emotion of the video, which they also grouped into 27 categories. [22]

arousal. Valence captures the positivity or negativity of an affect while arousal captures the intensity of the affect [66]. Figure 4 shows this model, and how different emotion keywords were represented on it. The Circumplex Model has been successfully applied to other languages and cultures [70].

Watson and Tellegen introduced the Positive-Affect Negative-Affect (PANA) model for mood with two dimensions of positive and negative affect, shown in figure 5 [77]. The PANA model is similar to a rotated Circumplex Model.

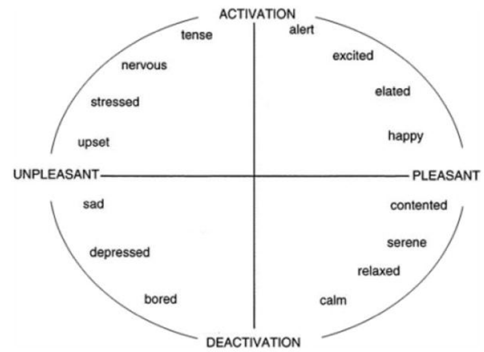

Figure 4. The Circumplex Model describes emotions in two dimensions: valence (x-axis) and arousal (y-axis). [62]

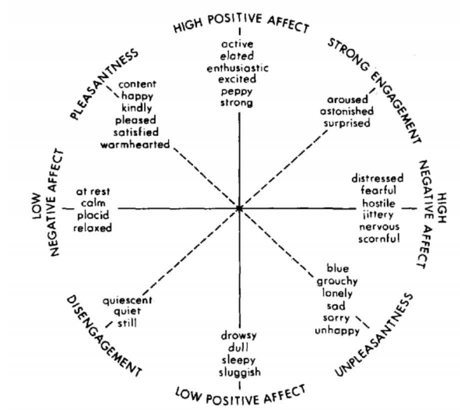

Figure 5. The PANA Model is a 2-dimensional model with the x-axis representing the level of negative affect and the y-axis representing the level of positive-affect. [77]

Mehrabian's Pleasure, Arousal and Dominance (PAD) model is a 3-dimensional emotion model [56]. The first two dimensions are similar to the Circumplex Model, but the added third dimension represents the dominance or submissiveness of an emotion. For example, fear and anger both represent emotions with a low valence and a high arousal, but fear is a more submissive emotion and anger is a more dominant emotion [56].

All emotion models share a similar goal of providing a representation for a set of emotions. The choice of emotion model is largely a product of the specific goals of a research experiment. Categorical models can be easier to label than dimensional models because words are often more intuitive than numbers when conceptualizing emotion. Dimensional models have the advantage of being able to express subtle changes in emotion more effectively because they utilize real values. Understanding the limitations of an emotion model and experimenting with other models will lead to the development of more robust automated emotion perception systems.

4. Facial Expressions and Perception

Facial expressions are the configurations created by the movement of muscles in the face. They are known to play a vital role in nonverbal communication and widely believed to convey emotional information [2]. Out of 149 scientists, 80% believe that there are universal signals of emotion exhibited in the face or voice [30], but the extent to which facial expressions are linked to an underlying emotion is not fully understood. In this section we provide an overview of the relationship between facial expressions and emotions. For a more thorough survey we recommend [8].

4.1. Encoding the Face

Ekman and Friesen created the Facial Action Coding System (FACS) in 1978, updated in 2002, which has become one of the most ubiquitous systems for encoding facial features [36,75]. FACS is based on mapping facial muscle movements to a set of action units that represent nearly all possible facial movements. Figure 6 shows examples of commonly used action units in FACS. Each of these action units are also paired with an intensity value that represents how present any given action unit is in a face, measured on a scale of A-E, where E is the most intense. Various studies have verified the reliability of FACS measures [21, 71].

Figure 6. Some common Action Units in the Facial Action Coding System, along with a visual example and description. [11]

4.2. Posed and Spontaneous Facial Expressions

Posed facial expressions are expressions that are deliberately created, such as a smile for a photograph. Spontaneous facial expressions, on the other hand, are created in response to some stimulus, such as a smile that occurs after someone hears a funny joke. Smiles are the most well-studied expression when comparing posed versus genuine, with genuine smiles being commonly referred to as Duchenne smiles [40]. The presence of FACS action unit 6 is a distinguishing characteristic of Duchenne smiles, as shown in figure 7 [13].

Figure 7. Comparison of Duchenne and Non-Duchenne Smiles from two people. Duchenne smiles exhibit AU6, cheek raiser, while Non-Duchenne smiles do not. The letters A-E represent the intensity of the AU. [13]

There are key differences between posed and spontaneous expressions [8]. [64] hypothesizes that posed expressions more closely resemble stereotypical ideas about facial expressions than spontaneous expressions. There is some evidence for voluntary and involuntary facial expressions being linked to different neural circuits, which may affect their expression [15, 63, 65].

Posed and spontaneous facial expressions are distinct in both their creation method and resulting muscular configuration. Spontaneous expressions have been shown to be more difficult for people to match with emotion keywords than posed expressions [57]. The use of posed expressions in emotion studies is criticized because of the bias it can introduce towards emotions that more closely match our preconceptions of what emotion expressions should look like [8]. Despite the differences in posed and spontaneous expressions, [42] found that people do poorly at identifying an expression as posed or spontaneous, with people rating expressions as spontaneous more often than they truly are.

4.3. Microexpressions

Microexpressions are generally regarded as brief facial expressions that occur when people are trying to conceal true emotions [34, 43]. Duration is a key defining aspect of a microexpression. Microexpressions have a duration of 170-500ms, similar to that of a blink [78]. The other defining aspect of microexpressions is that they occur when an emotion's expression is inhibited, and there is evidence to support this behavior [35, 61]. However, microexpressions have not been found to consistently exist when someone is concealing their emotion [61].

Microexpressions have been touted for their ability to detect deception in popular culture with the 2009 TV series Lie to Me, and they have been incorporated into law enforcement training through the Wizards project [14]. However, there have been very few empirical studies of microexpressions. And results from studies in microexpression research do not provide much support for the ability of microexpressions to reveal deception.

[55] claims to find the first systematic evidence for the ability of microexpressions to differentiate liars and truthtellers [55], but [18] and [61] find that short, involuntary expressions occur in both liars and truthtellers, and that the findings are so inconsistent that conclusions should not be drawn from these involuntary expressions. Overall, the body of work on microexpressions is small and there is still more evidence needed for the creation, duration, and form of microexpressions.

4.4. Perception and Culture

Supporters of the universality hypothesis believe that some facial expressions are universal indicators of certain emotions, independent of other factors like culture [33]. Others argue that contextual factors play a large role in emotion perception [10].

In 1971, Ekman and Friesen performed a seminal study in New Guinea that provided evidence of a universal relationship between facial expressions and emotions [33]. The study identified a group of individuals with minimal exposure to Western culture. A translator told the group an emotional story, and the participants were asked to choose the picture which best represented the emotion in the story from a set with various facial expressions. The researchers found high percentages of respondents choosing the face that matched the intended emotion of the story. Ekman et al. performed a similar experiment comparing judgments of emotions and their perceived intensity across 10 cultures, and found consensus on the classification of emotion and classification of relative intensity from the facial expression picture [32].

Some argue that a forced-choice format, within-subjects test design, and use of posed expressions have biased results in favor of universality [69]. Gendron et al. compares the results of a remote culture and a Western culture tasked with sorting facial expression pictures into their own clusters of emotion categories, and found that the clusters did not follow a universal pattern [38]. Jack et al. showed animations to two groups of different cultures, and found that the mental representations associated with emotion categories differ according to culture [46].

4.5. Perception and Context

In addition to culture, there is also research investigating the importance of context in perceiving emotion from facial expressions. This context can include location, background information, body, and any voice. Figure 8 illustrates the effect that context can have on emotion perception by showing Serena William's facial expression with and without context [10]. Without context, it is possible to perceive anger or pain from the face. But with context, it is much clearer that she is overjoyed and triumphant.

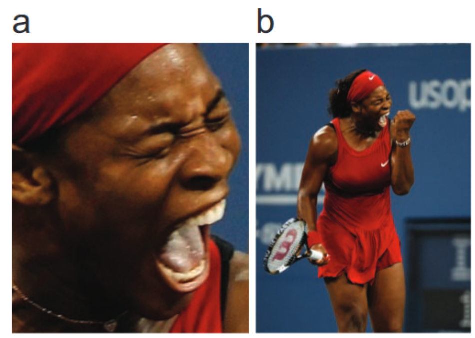

Figure 8. Comparison of Serena William's expression with and without context [10]: (a) without context can signal various emotions and (b) with context is very likely to signal joy.

People are extremely limited in recognizing emotion without context, highlighting the crucial role that context plays in emotion perception. [41] finds that configurations of facial muscles are ambiguous in determining emotion, and we recognize emotions in face-context combinations. [9] finds that when people are asked to recognize emotion, they better remember the context, which provides evidence that context is encoded in memory when an emotion is being perceived. The effect of context has been shown to be dependent on culture, as Japanese subjects were found to look at surrounding expressions when perceiving emotion, while Western subjects were more likely to look at an individual's expression [54]. Contextual factors are also considered to affect emotion perception on an individual level, as it has been found that each individual labeler has their own patterns when labeling emotions [3].

5. Discussion

It is essential that advances in automated emotion perception take into account the nuances of human emotion perception to foster a better understanding of emotion and to create more robust systems built on solid scientific foundations. This work aims to align the computational and psychology domains of emotion science to have a common ground moving forward in automated and human emotion perception. In this section, we distill our findings and provide recommendations for future research in automated emotion perception.

5.1. Standards for Facial Expression Recognition

There is some ambiguity in the term "facial expression" as it used in emotion recognition, often being conflated with "emotion." Instead, a standard should be used, such as the Facial Action Coding System (FACS) for facial expression recognition. This prevents confusion between the terms "expression" and "emotion," and gives the problem of "facial expression recognition" a clear evaluation standard in action units. For example, CK+ provides FACS codings for facial expressions with additional emotion labels [52].

5.2. Emotion Labels as Evidence not Truth

Evidence links facial expressions, body language, situational context, cultural information, and information about the observer as factors in emotion perception. The exact role and influence that each factor has on emotion perception is not currently known. This more holistic view of emotion perception brings about important research questions regarding the relationship between facial expressions and emotions such as 1) "what facial expressions result in more consensus in emotion perception?", 2) "which expressions are more ambiguous without additional information?" and 3) "do the demographic characteristics of the observer impact the perceived emotion?" We propose that future research in automated emotion perception measure uncertainty in emotion labels, allow for multiple emotion labels for a given sample and collect observer information to identify patterns in responding.

5.3. Incorporate Culture and Context

Research highlights the importance of culture and context in human emotion perception [9, 41]. Data used for automated emotion perception fall into one of two categories: constrained (lab created or controlled) or unconstrained (real-world). While unconstrained data is preferred for real-world applications, there are many unanswered research questions as to the role of context in both human and automated emotion perception, which leaves constrained data as the best path forward to answer these questions. A structured approach to adding context would allow for proper evaluation of context in emotion perception.

While some work has been done on multi-modal emotion perception (e.g. a combination of audio, text and visual), most work to date focuses on improving performance using a single modality or simply combining all modalities with deep neural networks without identifying the factors that contribute to a particular decision. A first step in incorporating context is to utilize these datasets to pinpoint the individual and combined effects of audio, text and visual information on human and automated emotion perception.

In order to study the effects of context it is important to systematically add context to the data. For audio and text, adding context might simply involve introducing the audio or text from time-points before the current utterance. For visual, transitioning from the most to least restrictive setting might involve 1) a tightly cropped face making direct eye contact with the background removed, 2) adding changes in gaze, 3) adding movement, 4) adding background, 5) adding upper body, and finally 6) adding full body context. Combinations of the above contexts may also be used, resulting in a large number of settings.

Such a systematic and controlled study will require the collection of data in a lab using actors. While spontaneous expressions and unconstrained data are preferred, if we are to pinpoint the role of context in emotion perception, a certain level of precision is required. Otherwise, the field of automated emotion perception will continue on applying state-of-the-art deep learning models to the problem without any useful insights or true innovation.

With respect to the effect of culture on emotion perception, it is essential to tackle this problem with an interdisciplinary team with representatives from psychology, sociology, natural language processing and computer vision. A clear next step is to collect emotion perception data (videos and labels) from a wide variety of cultures and to maintain cultural information for both the data and the observers. With a large-scale dataset for this problem, significant analysis can be performed to begin to answer questions about the effect of culture on the presentation and perception of emotion. As with the context studies, it is essential to approach this problem systematically and in very controlled settings in order to isolate the variables of interest (i.e. the culture of the individuals presenting the emotion and of the individuals perceiving the emotion).

5.4. Posed vs. Spontaneous Emotion Expressions

Posed and spontaneous facial expressions are different in both appearance and in the context that they are created [8, 13]. However, humans are not very good at determining the difference between posed and spontaneous expressions [42]. The elicitation of an expression is a key advantage of many lab-controlled datasets where that information is known. Unconstrained datasets (often collected from the internet using keyword search) do not discern between posed and spontaneous expressions, which can be problematic due to their inherent differences. We suggest that researchers avoid the use of web-collected data for the problem of emotion perception as the conditions under which the data was collected (e.g. posed vs. spontaneous, context and cultural information, etc.) are unknown.

When trying to perceive emotion, spontaneous expressions are likely to be better indicators of emotion because they are often elicited by an emotional response. Still, spontaneous expression datasets have constraints that are not present in the real world. One example of this is that many lab-controlled elicitations of emotion involve the subject starting from a neutral state. As emotion perception is essential in everyday interactions, data that represents everyday interactions (i.e. conversational dyads and groups) should be used for the study of human and automated emotion perception. We propose that future researchers collect data in controlled environments with more realistic scenarios. Specifically, using egocentric cameras to collect conversational data from the perspective of each participant will allow for the analysis of emotion elicitation and perception in real world situations without adding additional confounding factors from completely unconstrained data.

5.5. Subtle Expressions over Microexpressions

There is very little research investigating microexpression, despite their prevalence in popular culture [18]. Additionally, there is little evidence to support their nature, their implications, and even if they are a useful concept for human emotion perception. We recommend research in automated emotion perception explore paths that have stronger scientific foundations than microexpressions which have relatively little empirical validation. Specifically, many automated approaches perform better with exaggerated, posed expressions than they do with subtle, spontaneous expressions [50]. We propose that future work focus on detecting the subtle expressions humans exhibit in real-world conditions rather than microexpressions. As we encourage the use of conversational scenarios for their ability to produce spontaneous expressions, we also encourage them for their ability to produce subtle expressions.

5.6. Interdisciplinary Approaches

It can be challenging for researchers in automated emotion perception to have a deep understanding of human emotion perception, and psychologists have been critical of work in automated emotion perception, highlighting the importance of understanding human emotion perception [8]. Automated emotion perception is a field which requires perspectives from psychology, sociology, natural language processing and computer vision at the very least. Previous automated emotion perception works that have been created with interdisciplinary perspectives tend to be more aligned with human emotion perception research than works lacking that perspective. For example, the CK+ dataset was created as a collaborative effort between researchers with both a psychology and computer science background, and they were careful to describe the limitations of emotion labels and only accept labels as valid if they met specific criteria [52]. We should pursue interdisciplinary study in this area as it will improve the quality of research being done and will advance the fields of automated and human emotion perception. As a crucial next step, we propose that interdisciplinary workshops be held to bridge the gap between research in automated and human emotion perception.

6. Conclusion

Emotion perception is a challenging field to study as it investigates complex human behaviors that are not fully understood. This work provides an interdisciplinary discussion of automated and human emotion perception. The goal of this work is to align both perspectives (psychology and computer science, i.e. human and automated) of emotion perception, facilitating future discussions and research in this interdisciplinary area.

Human emotion perception research is shifting away from the idea that emotions exist in a discrete, biologically-based set, and instead the concept might be constructed based on a human experience with emotion that is shaped by both natural and environmental factors. The extent to which facial expressions reflect emotion is still being studied, with the theory that they are universally linked being questioned by evidence of culture and contextual factors impacting emotion perception.

Current automated emotion perception systems can detect stereotypical, exaggerated expressions and assign an emotion from this, but tend to struggle to perceive emotion from more subtle and realistic examples of emotion. The underlying assumption that an emotion can always be perceived from a facial expression is not valid, and emotion perception systems must grow past this idea – ideally using multi-modal data – to achieve better performance in real-world scenarios.

We provide a series of recommendations for automated emotion perception based on some of the disparities we have identified between automated and human emotion perception research. In our discussion, we focus on identifying a standard for facial expressions, quantifying uncertainty in emotion labels, systematically incorporating culture and context, understanding limitations associated with posed and spontaneous emotion expressions, utilizing subtle expressions in place of microexpressions and finally approaching the problem of emotion perception as an interdisciplinary one incorporating perspectives from psychology, sociology, natural language processing and computer vision.

It will be challenging to construct datasets and develop methods that begin to capture the complexities of human emotion perception. However, it is vital that a modern social science perspective be internalized by researchers in automated emotion perception. This perspective can greatly improve the capacity for machines to perceive emotions as we humans do naturally, and ultimately produce higher quality research than is possible without an understanding of human emotion perception.

References

[1] *The nature of emotion : fundamental questions*. Series in affective science. Oxford University Press, New York, 1994. 1

[2] Ralph Adolphs. Social cognition and the human brain. *Trends in Cognitive Sciences*, 3(12):469–479, 1999. 1, 5

[3] Hillel Aviezer, Noga Ensenberg, and Ran R Hassin. The inherently contextualized nature of facial emotion perception. *Current Opinion in Psychology*, 17:47–54, 2017. Emotion. 6

[4] Lisa Barrett and Tor Wager. The structure of emotion: Evidence from neuroimaging studies. *Current Directions in Psychological Science*, 15:79–83, 04 2006. 2

[5] Lisa Feldman Barrett. Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, 10(1):20–46, 2006. PMID: 16430327. 2

[6] Lisa Feldman Barrett. The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(1):1–23, 10 2016. 2

[7] L. F. Barrett. *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt, 2017. 2

[8] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1):1–68, 2019. PMID: 31313636. 5, 7, 8

[9] Lisa Feldman Barrett and Elizabeth A. Kensinger. Context is routinely encoded during emotion perception. *Psychological Science*, 21(4):595–599, 2010. PMID: 20424107. 6, 7

[10] Lisa Feldman Barrett, Batja Mesquita, and Maria Gendron. Context in emotion perception. *Current Directions in Psychological Science*, 20(5):286–290, 2011. 6

[11] Delina Beh Mei Yin, Shariman Omar, Bazilah A. Talip, Amalia Mukhlas, Nur Norain, and Abu Talib Othman. Fusion of face recognition and facial expression detection for authentication: a proposed model. pages 1–8, 01 2017. 5

[12] Igor Bisio, Alessandro Delfino, Fabio Lavagetto, and Mario Marchese. *Opportunistic Detection Methods for Emotion-Aware Smartphone Applications*, pages 53–85. 11 2013. 4

[13] Yevgen Bogodistov and Florian Dost. Proximity begins with a smile, but which one? associating non-duchenne smiles with higher psychological distance. *Frontiers in Psychology*, 8:1374, 2017. 5, 7

[14] Charles F Bond Jr and Ahmet Uysal. On lie detection "wizards". *Law and human behavior*, 31(1):109–115, 2007. 6

[15] Joan C. Borod, Elissa Koff, and Betsy White. Facial asymmetry in posed and spontaneous expressions of emotion. *Brain and cognition*, 2(2):165–175, 1983. 5

[16] Tobias Brosch, Klaus Scherer, Didier Grandjean, and David Sander. The impact of emotion on perception, attention, memory, and decision-making. *Swiss medical weekly*, 143:0, 05 2013. 1

[17] Aaron Frederick Bulagang, Ng Giap Weng, James Mountstephens, and Jason Teo. A review of recent approaches for emotion classification using electrocardiography and electrodermography signals. *Informatics in Medicine Unlocked*, 20:100363, 2020. 2

[18] Judee Burgoon. Microexpressions are not the best way to catch a liar. *Frontiers in Psychology*, 9, 09 2018. 6, 8

[19] A. Celeghin, M. Diano, Arianna Bagnis, M. Viola, and M. Tamietto. Basic emotions in human neuroscience: Neuroimaging and beyond. *Frontiers in Psychology*, 8, 2017. 2

[20] Alessia Celeghin, Matteo Diano, Arianna Bagnis, Marco Viola, and Marco Tamietto. Basic emotions in human neuroscience: Neuroimaging and beyond. *Frontiers in Psychology*, 8:1432, 2017. 2

[21] Jeffrey Cohn, Zara Ambadar, and Paul Ekman. Observer-based measurement of facial expression with the facial action coding system. *The Handbook of Emotion Elicitation and Assessment*, 01 2007. 5

[22] Alan S. Cowen and Dacher Keltner. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, 114(38):E7900–E7909, 2017. 3, 4

[23] Charles Darwin. *The Expression of the Emotions in Man and Animals*. Cambridge Library Collection - Darwin, Evolution and Genetics. Cambridge University Press, 2 edition, 2009. 2

[24] John Dewey. The theory of emotion: I: Emotional attitudes. *Psychological Review*, 1(6):553–569, 1894. 2

[25] Shichuan Du, Yong Tao, and Aleix M. Martinez. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, 111(15):E1454–E1462, 2014. 3

[26] Knight Dunlap. Are emotions teleological constructs? *The American journal of psychology*, 44(3):572–576, 1932. 2

[27] P Ekman. California mental health research digest. *Naval Research Logistics Quarterly*, 1970. 3

[28] P. Ekman. An argument for basic emotions. *Cognition & Emotion*, 6:169–200, 1992. 2

[29] Paul Ekman. *Basic Emotions*, chapter 3, pages 45–60. John Wiley & Sons, Ltd, 1999. 3

[30] Paul Ekman. What scientists who study emotion agree about. *Perspectives on Psychological Science*, 11(1):31–34, 2016. PMID: 26817724. 5

[31] P. Ekman and W. Friesen. A new pan-cultural facial expression of emotion. *Motivation and Emotion*, 10:159–168, 1986. 3

[32] Paul Ekman, Wallace Friesen, Maureen O'Sullivan, A. Chan, Irene Diacoyanni-Tarlatzis, Karl Heider, Rainer Krause, William LeCompte, Tom Pitcairn, and Pio Ricci Bitti. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53:712–7, 11 1987. 6

[33] Paul Ekman and Wallace V. Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2):124–129, 1971. 2, 6

[34] Paul Ekman and Wallace V. Friesen. Detecting deception from the body or face. *Journal of personality and social psychology*, 29(3):288–298, 1974. 5

[35] Paul Ekman, E. T. Rolls, D. I. Perrett, and H. D. Ellis. Facial expressions of emotion: An old controversy and new findings [and discussion]. *Philosophical transactions. Biological sciences*, 335(1273):63–69, 1992. 6

[36] Hager JC Ekman P, Friesen WV. *Facial Action Coding System: The Manual on CD ROM*. A Human Face, 2002. 5

[37] M. Gendron and Lisa Feldman Barrett. Reconstructing the past: A century of ideas about emotion in psychology. *Emotion Review*, 1:316 – 339, 2009. 1, 2

[38] M. Gendron, D. Roberson, Jacoba Marietta van der Vyver, and L. F. Barrett. Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion*, 14 2:251–62, 2014. 6

[39] L Greenberg and Jeremy Safran. Emotion in psychotherapy. *The American psychologist*, 44:19–29, 02 1989. 1

[40] Sarah D. Gunnery and Mollie A. Ruben. Perceptions of duchenne and non-duchenne smiles: A meta-analysis. *Cognition and Emotion*, 30(3):501–515, 2016. PMID: 25787714. 5

[41] Ran R. Hassin, Hillel Aviezer, and Shlomo Bentin. Inherently ambiguous: Facial expressions of emotions, in context. *Emotion Review*, 5(1):60–65, 2013. 6, 7

[42] Ursula Hess and Robert E. Kleck. The cues decoders use in attempting to differentiate emotion-elicited and posed facial expressions. *European journal of social psychology*, 24(3):367–381, 1994. 5, 7

[43] Carolyn M. Hurley, Ashley E. Anker, Mark G. Frank, David Matsumoto, and Hyisung C. Hwang. Background factors predicting accuracy and improvement in micro expression recognition. *Motivation and emotion*, 38(5):700–714, 2014. 5

[44] Daniel D Hutto, Ian Robertson, and Michael D Kirchhoff. A new, better bet: Rescuing and revising basic emotion theory. *Frontiers in psychology*, 9:1217–1217, 2018. 2

[45] Carroll E. Izard. *Human emotions / [edited] by Carroll E. Izard*. Plenum Press New York, 1977. 2

[46] Rachael E. Jack, Oliver G. B. Garrod, Hui Yu, Roberto Caldara, and Philippe G. Schyns. Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America*, 109(19):7241–7244, 2012. 6

[47] William James. What is an emotion? *Mind*, 9(34):188–205, 1884. 2

[48] Dacher Keltner, Disa Sauter, Jessica Tracy, and Alan Cowen. Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior*, 43, 06 2019. 2

[49] Jyoti Kumari, R. Rajesh, and K.M. Pooja. Facial expression recognition: A survey. *Procedia Computer Science*, 58:486–491, 2015. Second International Symposium on Computer Vision and the Internet (VisionNet'15). 1

[50] Shan Li and Weihong Deng. Deep facial expression recognition: A survey. *IEEE transactions on affective computing*, pages 1–1, 2020. 1, 8

[51] Kristen A Lindquist and Lisa Feldman Barrett. Constructing emotion: The experience of fear as a conceptual act. *Psychological science*, 19(9):898–903, 2008. 2

[52] Patrick Lucey, Jeffrey F Cohn, T Kanade, Jason Saragih, Zara Ambadar, and I Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pages 94–101. IEEE, 2010. 7, 8

[53] George Mandler. William james and the construction of emotion. *Psychological science*, 1(3):179–180, 1990. 2

[54] Takahiko Masuda, Phoebe Ellsworth, Batja Mesquita, Janxin Leu, Shigehito Tanida, and Ellen Veerdonk. Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*, 94:365–381, 04 2008. 6

[55] David Matsumoto and Hyisung C. Hwang. Microexpressions differentiate truths from lies about future malicious intent. *Frontiers in Psychology*, 9:2545, 2018. 6

[56] Albert Mehrabian. *Basic dimensions for a general psychological theory: implications for personality, social, environmental, and developmental studies*. Oelgeschlager, Gunn and Hain, Cambridge, 1980. 3, 4

[57] Michael T. Motley and Carl T. Camden. Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western journal of speech communication*, 52(1):1–22, 1988. 5

[58] W. G. Parrott. *Emotions in social psychology: essential readings*. Psychology Press, Hove [England];Philadelphia;, 2001. 3

[59] Robert Plutchik. A psychoevolutionary theory of emotions. *Social Science Information*, 21(4-5):529–553, 1982. 3

[60] Robert Plutchik. The nature of emotions. *American scientist*, 89(4):344–344, 2001. 3

[61] Stephen Porter and Leanne ten Brinke. Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science*, 19(5):508–514, 2008. PMID: 18466413. 6

[62] J. Posner, J. Russell, and B. Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17:715 – 734, 2005. 4

[63] William E. Rinn. The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological bulletin*, 95(1):52–77, 1984. 5

[64] Michael D. Robinson and Gerald L. Clore. Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological bulletin*, 128(6):934–960, 2002. 5

[65] Elliott D. Ross and Vinay K. Pulusu. Posed versus spontaneous facial expressions are modulated by opposite cerebral hemispheres. *Cortex*, 49(5):1280–1291, 2013. 5

[66] James Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 12 1980. 3, 4

[67] J. Russell. Core affect and the psychological construction of emotion. *Psychological review*, 110 1:145–72, 2003. 2

[68] James Russell and Lisa Barrett. Core affect, prototypical emotional episodes, and other things called emotion: Dis-

secting the elephant. *Journal of personality and social psychology*, 76:805–19, 06 1999. 2

[69] J. A. RUSSELL. Is there universal recognition of emotion from facial expression ? a review of the cross-cultural studies. *Psychological bulletin*, 115(1):102–141, 1994. 1, 6

[70] James A. Russell, Maria Lewicka, and Toomas Niit. A cross-cultural study of a circumplex model of affect. *Journal of personality and social psychology*, 57(5):848–856, 1989. 4

[71] M.A. Sayette, Jeffrey Cohn, J.M. Wertz, M.A. Perrott, and D.J. Parrott. A psychometric evaluation of the facial action coding system for assessing spontaneous expression. *Journal of Nonverbal Behavior*, 25, 01 2001. 5

[72] Klaus R. Scherer, Elizabeth Clark-Polner, and Marcello Mortillaro. In the eye of the beholder? universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, 46(6):401–435, 2011. 2

[73] Paul Silvia. Emotional responses to art: From collation and arousal to cognition and emotion. *Review of General Psychology*, 9:342–357, 12 2005. 1

[74] Silvan S. Tomkins. *Affect, imagery, consciousness: Vol. I. The positive affects.* Affect, imagery, consciousness: Vol. I. The positive affects. Springer, Oxford, England, 1962. 2

[75] Ekman P. & Friesen W. V. *The Facial Action Coding System*. Consulting Psychological Press, 1978. 5

[76] G. R. VandenBos. *APA Dictionary of Psychology*. American Psychological Association, 2007. 2

[77] David Watson and Auke Tellegen. Toward a consensual structure of mood. *Psychological bulletin*, 98(2):219–235, 1985. 3, 4

[78] Wen-Jing Yan, Qi Wu, Yu-Hsin Chen, Jing Liang, and Xiaolan Fu. How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of Nonverbal Behavior*, 37, 12 2013. 5