# Context in Automated Affect Recognition

**Matthew Groh**
MIT Media Lab
Cambridge, MA 02139
groh@mit.edu

**Rosalind Picard**
MIT Media Lab
Cambridge, MA 02139
picard@media.mit.edu

## Abstract

Affect recognition depends on interpreting both expressions and their associated context. While expressions can be explicitly measured with sensor technologies, the role of context is more difficult to measure because context is often left undefined. In an effort to explicitly incorporate pragmatics in automated affect recognition, we develop a framework for categorizing context. Building upon ontologies in affective science and symbolic artificial intelligence, we highlight seven key categories: ambient sensory environment, methods of measurement, semantic representation, situational constraints, temporal dynamics, sociocultural dimensions, and personalization. In this short paper, we focus on how the epistemological categories of context influence the training and evaluation of machine learning models for affect recognition. Incorporating context in the practical and theoretical development of affect recognition models is an important step to developing more precise and accurate models.

## Context Affects Meaning

In an early 20th century film experiment, cinematographer Lev Kuleshov presented audiences with a short clip of an actor expressing a neutral facial expression followed by one of three scenes: a bowl of soup, a young girl in a coffin, and a woman lying on a couch. Depending on which scene the audience saw, the audience described the actor's expression as indicative of different emotions; hunger for the soup, sadness for the deceased, and lust for the woman. Two recent experiments replicated the results of the original Kuleshov experiment and extended it to show that scenes conveying fear and desire also lead audiences to report neutral facial expressions as expressions matching the sentiment in the juxtaposed scenes [3, 8].

Context shapes how humans perceive and recognize emotions. For example, the art of transforming a script into a heart-wrenching movie involves not only actors' dialogue and physicality (their observable expressions) but also how these expressions relate to scene transitions, the musical accompaniment, lighting conditions, costume and set design, and narrative devices. Likewise, how an observer interprets another person's smile depends on contextual cues like whether a person is acting earnestly, whether a person is in a pain-eliciting situation, or whether social display rules might influence a person to mask their inner feelings.

In affective computing, emotion recognition has been described as a combination of "observations of emotional expressions" and "reasoning about an emotion-generating situation" [40]. This dual focus of emotion recognition on expressions and context matches research in affective science, which shows observable expressions are often ambiguous without context [1, 2, 4, 7, 15, 18, 22, 23, 24, 28, 39, 47, 48]. Emotion recognition is a subset of affect recognition, which is sometimes referred to as affect detection, affect estimation, and affect measurement in the field of affective computing. Emotion recognition has also been called empathic accuracy in the field of affective science and emotion reasoning in the field of developmental psychology [30, 41]. Automated affect recognition

Meaning in Context: Pragmatic Communication in Humans and Machines Workshop at 35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

applies methods from signal processing and machine learning to situated expression data, which are data on observable expressions and their associated context [19]. While facial expressions, physical gestures, speech prosody, physiology (heart-rate, breathing-rate, and electrodermal activity), and other human behavior are all concrete examples of observable expressions, context is more amorphous and generally refers to the relationship of these expressions to each other and the external environment [10]. Moreover, context is multidimensional and difficult to circumscribe with a single label. In a recent experiment examining facial expressions across contexts in video, context is defined as the 653 categories that a neural network has been trained to classify, which include categories such as breakfast, car, humor, airport, lake, bottle, and mother where mother can refer or "pertain to mothers in any number of ways, ranging from footage of actual parenting to a man discussing his mother" [12]. This definition describes an algorithmic classification schema that identifies potentially useful yet vague aspects of context.

## Building a Framework for Context

How can we systematically identify the roles of context in automated affect recognition? First, we need a language to discuss what we mean by context in affect recognition. In the abstract, context represents a complex high-dimensional feature space representing the inter-relatedness of elements that are often only partially available to observers. In affective science, context has been described as the collective "unmeasured factors" that contribute to how emotions are constructed; in the same paper, the authors describe the most salient, yet often unmeasured contexts as situational, social, physical, mental, temporal, personal, and cultural [6]. Likewise, in another review of context in emotion research, context is presented as a framework made up by three major components: personal, situational, and cultural features [25]. By explicitly identifying these categories rather than leaving context as a catch-all term for anything unmeasured, we can begin to build a framework to more precisely evaluate and discuss the varying roles of context in affect recognition.

When we examine affect recognition performed by computers, we need to take additional context into account. In the field of symbolic artificial intelligence (AI), ontology engineers have developed frameworks for incorporating context in common sense reasoning on natural language processing tasks [26, 36]. These frameworks have been useful for identifying assumptions that are often taken for granted in human communication but necessary for machine communication. In particular, the context identified in symbolic AI includes epistemological components addressing system-level questions like how to arbitrate opposing perspectives, what serves as evidence, what can be assumed, what expertise is required for making observations and judgments, and who believes a claim and why. In affect recognition within the context of affective computing, these questions become: how do we semantically represent affect, how do we label data, what do we assume about the accuracy of any human or machine appraisal of affect, what qualifies someone to label data, and how do we evaluate a model's performance.

Drawing from and expanding on entry points from both affective science and symbolic AI, we identify seven key categories to consider in automated affect recognition: ambient sensory environment, methods of measurement, semantic representation, situational constraints, temporal dynamics, sociocultural perspectives, and personalization. Our aim in establishing this seven-category framework is not to establish a new theory of emotions nor to claim there cannot be an eighth category, but instead, our aim is to take concrete steps toward unifying the many useful elements of context for affect recognition that have been already articulated in the affective science and affective computing literature. As such, we aim to synthesize a framework that provides both a theoretical foundation and a practical set of constructs. We are most inspired when theory and practice support each other, and since practice in affect recognition is growing rapidly, we seek to advance a theoretical framework for context that can grow with it, supporting and strengthening the growing practice. We describe the seven categories briefly below.

The first category, ambient sensory environment, refers to the sensory aspects of one's immediate surrounding settings e.g., the weather, soundscape, scenery, and smells. While ambient sensory environment does not neatly fit into any of the categories specified by Lenat 1998 or Barrett et al 2019, ambient sensory environment includes the face-context pairings described in earlier affective science research e.g., "face imbedding" (information within an image around a target face) and "response coherence" (information on congruence of facial expressions with non-facial expressions) [38]. The next two categories, methods of measurement and semantic representation are based on the five categories in the symbolic AI framework which focus on the system-level, epistemological concerns that are relevant for training machine learning models to predict affect labels. The final four categories occur in both Lenat 1998 and Barrett et al 2019. Situational constraints refer to constraints imposed by the activity or venue within which something is happening. For example, the inability to safely take one's eyes off the road while driving is a situational constraint. In the Lenat 1998 framework, situational constraints are further divided into topic/usage (referring to activity) and absolute place (referring to a place like the pyramids of Giza or the Golden Gate bridge) and type of place (referring to a place like a pizza joint or a shower), and here, we address all three of these categories together. Temporal dynamics refer to the dynamic nature of expressions and the trajectory and seasonality of emotional events. Sociocultural dimensions are the components of context related to other people. Finally, personalization refers to individuals' idiosyncrasies, which can range from an individuals' tastes and preferences to mental disabilities. All of these categories of context can overlap, and they are not mutually exclusive. These categories serve as a starting point to systematically examine how each different component of context situates expressions and shapes the appraisal and recognition of emotions.

## Evaluating Automated Affect Recognition

Instead of detailing the role of each category here, we address the epistemological categories (methods of measurement and semantic representation) by asking: how can we evaluate the accuracy of an affect recognition model? In order to empirically evaluate a statistical learning model, we identify a source of human-provided (ground truth) labels, $y$, upon which to compare the model's predictions, $\hat{y}$. For affect recognition, ground truth labels usually come from one or more of these sources: individuals' self-reports of what they feel, external observers' reports of what they perceive others to experience, and experimentally-elicited or situationally-driven emotions. These three different methods of measurement are all useful yet imperfect for representing ground truth.

Self-reports provide an opportunity to collect ground truth labels based on an individual's inner feelings, but self-reports are subject to willful deception, can be inhibited by interoceptive ability and alexithymia, and are subject to social and cognitive biases [31]. For example, acquiescence bias is one particularly pernicious bias where research participants tend to agree with what they think the researchers want to hear [45].

External observers' reports can be collected by impartial and emotionally intelligent third parties. Most adult human observers know that outward appearance of affect does not necessarily reflect an individual's inner feelings, and as such, observation generally involves applying theory of mind and pragmatic reasoning about the target individual's expressions and situation before assigning an emotion label. Nonetheless, external observers' reports (just like self-reports) are not guaranteed to match an individuals' inner feelings. Moreover, while this approach allays concerns about willful deception and interoceptive ability, it cannot rule out social and behavioral biases of observers. One advantage of examining external observers' labels (as opposed to self-reports) is the ability to control the information to which the observers have access (e.g., a video with audio, audio only, silent video, a video with a mask over the target individual or background, a full body photograph, a photograph showing only the face, or many other permutations), because manipulation of information modalities enables research into context effects.

The third approach to collecting ground truth labels is generating situations known to elicit particular emotions. For example, an experiment could elicit affect by asking a participant to reflect on a past emotional experience, ask a participant to count backwards from 100 by 7s (which often elicits stress), or routing a participant's car into rather than away from traffic jams (which often elicits stress or sometimes anger) [33, 37]. However, experiments designed to elicit emotions in participants do no always elicit the intended emotions because people respond to different situations differently. Moreover, laboratory conditions often do not match real-world settings, which raises questions about how well the findings of an experiment generalize to the real-world.

Measuring affect requires selecting a method for representing affect. It is well known that affect can be represented as: continuous affective dimensions (e.g., valence, arousal, dominance) and discrete emotion categories (e.g., joy, anger, fear, sadness, disgust, surprise). Affect can also be represented as: emotion categories connected by continuous gradients (e.g., horror, fear, disgust, anxiety), mixtures of emotion categories (e.g., angrily surprised, sadly fearful), enduring states (e.g., frustration, stress, pain, anxiety, depression), or even by sets of symbols like emojis, which can represent discrete or mixed and overlapping states. For example, emojis can represent emotions that are otherwise difficult to express via text. By training a machine learning model on a large corpus of tweets using sets of emojis as labels, researchers achieved state-of-the-art performance on three natural language processing benchmark tasks including emotion classification, sentiment analysis, and sarcasm detection [21]. While there are many competing theories of emotion, there is no universal agreement on how emotion should be represented [5, 13, 14, 17, 20, 29, 32, 34, 42, 43, 44]. The choice of how affect is represented will influence how an affect recognition model is trained and ultimately how accurately it recognizes affective states.

We evaluate the accuracy of an affect recognition model and its generalizability on data the model has never previously seen. Consider a model represented algebraically as $\hat{y} = f(x, c)$ where $\hat{y}$ represents the predicted affect label, $x$ indicates the physical expression data, and $c$ signifies context. Once the model has been trained on an initial dataset, we can evaluate its performance on a hold-out set and compare $\hat{y}$, the machine-predicted affect labels, to $y$, the human-provided labels. This allows us to evaluate a range of accuracy metrics including sensitivity, specificity, F1-score, AUC, log-loss, Pearson correlation coefficient, Matthew's correlation coefficient, and Cohen's kappa among others. In assessing how well a model recognizes self-reported emotions or observed emotional states, "a reasonable criterion of success is to get a computer to recognize affect as well as another person, i.e., better than chance, but below 100% accuracy" [40]. In some instances where physiological signals from the autonomic nervous system are imperceptible to humans without computational tools, the evaluation criteria shift to how well these otherwise imperceptible signals predict experimentally elicited emotions or long-term measures of mental health like reduced stress, reduced casualties while driving, or better learning outcomes [11, 35].

In practice, the assumption that the training and holdout data are independent and identically distributed (i.i.d.) often does not hold because context changes. As such, real-world implementation of automated affect recognition systems needs to explicitly incorporate as much contextual information as possible to most effectively generalize and avoid spurious correlations between observable expressions and affective labels. A recent review on facial expressions and emotions concludes that context matters for interpreting emotions from facial expressions: "When facial movements do express emotional states, they are considerably more variable and depend on context," [6]. This conclusion refreshes the need to examine an engineering question: Can we measure the contexts that inform the relationship between facial expressions and emotional states? This is not a new question in the field of affective computing; the development of large-scale datasets for facial expression recognition in the wild (e.g., EmotiW, Aff-Wild) draws from the premise that context mediates how facial expressions are interpreted [16]. The limitations of context-free affect detection and the importance of context-awareness were discussed as core challenges to building affect recognition systems a decade ago [9, 27, 46].

Recent advances in sensing technology and neural networks have enabled researchers to incorporate context more effectively than ever before, which now raises additional questions: What contexts are informative for affect recognition, and how can we measure these contexts? In this short paper, we identify seven key categories of context that should be considered in artificial intelligence systems for affect recognition. In a future paper, we will review recent research in affective computing and affective science to demonstrate how the incorporation of these categories influence the accuracy of affect recognition systems.

## References

[1] Lior Abramson, Rotem Petranker, Inbal Marom, and Hillel Aviezer. Social interaction context shapes emotion recognition through body language, not facial expressions. *Emotion*, 2020.

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

[3] Daniel Barratt, Anna Cabak Rédei, Åse Innes-Ker, and Joost de Weijer. Does the Kuleshov effect really exist? Revisiting a classic film experiment on facial expressions and emotional contexts. *Perception*, 45(8):847–874, 2016.

[4] Lisa Feldman Barrett. Emotions are real. *Emotion*, 12(3):413, 2012.

[5] 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, 2017.

[6] 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.

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

[8] Marta Calbi, Katrin Heimann, Daniel Barratt, Francesca Siri, Maria A Umiltà, and Vittorio Gallese. How context influences our perception of emotional faces: A behavioral study on the Kuleshov effect. *Frontiers in psychology*, 8:1684, 2017.

[9] Rafael A. Calvo and Sidney D'Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1):18–37, 1 2010.

[10] Rafael A Calvo, Sidney D'Mello, Jonathan Matthew Gratch, and Arvid Kappas. *The Oxford handbook of affective computing*. Oxford University Press, USA, 2015.

[11] Yekta Said Can, Bert Arnrich, and Cem Ersoy. Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of biomedical informatics*, 92:103139, 2019.

[12] Alan S. Cowen, Dacher Keltner, Florian Schroff, Brendan Jou, Hartwig Adam, and Gautam Prasad. Sixteen facial expressions occur in similar contexts worldwide. *Nature*, 589(7841):251–257, 1 2021.

[13] Carlos Crivelli and Alan J Fridlund. Facial displays are tools for social influence. *Trends in Cognitive Sciences*, 22(5):388–399, 2018.

[14] Carlos Crivelli and Alan J Fridlund. Inside-out: From basic emotions theory to the behavioral ecology view. *Journal of Nonverbal Behavior*, 43(2):161–194, 2019.

[15] Beatrice de Gelder. Cultural differences in emotional expressions and body language. 2016.

[16] Abhinav Dhall. Context based facial expression analysis in the wild. In *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, pages 636–641, 2013.

[17] 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.

[18] Juan I Durán, Rainer Reisenzein, and José-Miguel Fernández-Dols. Coherence between emotions and facial expressions. *The science of facial expression*, pages 107–129, 2017.

[19] Sidney D'Mello, Arvid Kappas, and Jonathan Gratch. The affective computing approach to affect measurement. *Emotion Review*, 10(2):174–183, 2018.

[20] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[21] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*, 2017.

[22] José-Miguel Fernández-Dols and Carlos Crivelli. Emotion and expression: Naturalistic studies. *Emotion Review*, 5(1):24–29, 2013.

[23] Samuel W Fernberger. False Suggestion and the Piderit Model Author. Technical Report 4, 1928.

[24] Maria Gendron, Batja Mesquita, and Lisa Feldman Barrett. Emotion perception: Putting the face in context. 2013.

[25] Katharine H Greenaway, Elise K Kalokerinos, and Lisa A Williams. Context is everything (in emotion research). *Social and Personality Psychology Compass*, 12(6):e12393, 2018.

[26] Ramanathan V Guha. *Contexts: a formalization and some applications*. PhD thesis, Stanford University, 1992.

[27] Zakia Hammal and Merlin Teodosia Suarez. Towards context based affective computing. In *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, page 802, 2013.

[28] Ursula Hess and Shlomo Hareli. The social signal value of emotions: The role of contextual factors in social inferences drawn from emotion displays. 2017.

[29] Carl-Herman Hjortsjö. *Man's face and mimic language*. Studentlitteratur, 1969.

[30] William Ickes. Empathic Accuracy. *Journal of Personality*, 61:4, 1993.

[31] Daniel Kahneman and Alan B Krueger. Developments in the Measurement of Subjective Well-Being. Technical report, 2006.

[32] Dacher Keltner, Disa Sauter, Jessica Tracy, and Alan Cowen. Emotional expression: Advances in basic emotion theory. *Journal of nonverbal behavior*, pages 1–28, 2019.

[33] C. Kirschbaum, K. M. Pirke, and D. H. Hellhammer. The 'Trier social stress test' - A tool for investigating psychobiological stress responses in a laboratory setting. In *Neuropsychobiology*, volume 28, pages 76–81, 1993.

[34] Paul R Kleinginna and Anne M Kleinginna. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion*, 5(4):345–379, 1981.

[35] Rafal Kocielnik, Natalia Sidorova, Fabrizio Maria Maggi, Martin Ouwerkerk, and Joyce H D M Westerink. Smart technologies for long-term stress monitoring at work. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 53–58, 2013.

[36] Doug Lenat. The dimensions of context-space, 1998.

[37] Jennifer S Lerner, Ye Li, Piercarlo Valdesolo, and Karim S Kassam. Emotion and decision making. *Annual review of psychology*, 66:799–823, 2015.

[38] David Matsumoto and Hyi Sung Hwang. Judging Faces in Context. *Social and Personality Psychology Compass*, 4(6):393–402, 6 2010.

[39] Iris B Mauss and Michael D Robinson. Measures of emotion: A review. *Cognition and emotion*, 23(2):209–237, 2009.

[40] Rosalind W Picard. *Affective computing*. MIT press, 1997.

[41] Ashley L Ruba and Seth D Pollak. The Development of Emotion Reasoning in Infancy and Early Childhood. *The Annual Review of Developmental Psychology*, 2020.

[42] James A Russell. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145, 2003.

[43] James A Russell and Lisa Feldman Barrett. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, 76(5):805, 1999.

[44] Andrea Scarantino. How to do things with emotional expressions: The theory of affective pragmatics. *Psychological Inquiry*, 28(2-3):165–185, 2017.

[45] Peter B. Smith. Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, 35(1):50–61, 1 2004.

[46] Aggeliki Vlachostergiou, George Caridakis, and Stefanos Kollias. Investigating context awareness of Affective Computing systems: A critical approach. In *Procedia Computer Science*, volume 39, pages 91–98. Elsevier B.V., 2014.

[47] Matthias J Wieser and Tobias Brosch. Faces in context: a review and systematization of contextual influences on affective face processing. *Frontiers in psychology*, 3:471, 2012.

[48] Christine D Wilson-Mendenhall, Lisa Feldman Barrett, W Kyle Simmons, and Lawrence W Barsalou. Grounding emotion in situated conceptualization. *Neuropsychologia*, 49(5):1105–1127, 2011.