To appear in Emotion Review, Accepted for publication Jan 12, 2010.

Emotion research by the people, for the people

Rosalind W. Picard

MIT Media Laboratory

Cambridge, MA 02139

picard@media.mit.edu

Keywords: Affective computing, ambulatory monitoring, autonomic nervous system, idiographic research

Abstract:
Emotion research will leap forward when its focus changes from comparing averaged statistics of self-report data across people experiencing emotion in labs to characterizing patterns of data from individuals and clusters of similar individuals experiencing emotion in real life. Such an advance will come about through engineers and psychologists collaborating to create new ways for people to measure, share, analyze, and learn from objective emotional responses in situations that truly matter to people. This approach has the power to greatly advance the science of emotion while also providing personalized help to participants in the research.

Introduction

First, a story:

> A man in a colorful hot air-filled balloon realized he was lost. He spotted a woman digging in her garden and descended toward her, shouting, 'Excuse me, can you please help – tell me where I am? I promised a friend I would meet him an hour ago, and I'm lost. My friend will be upset and I must get there...
>
> The woman replied, 'You're in a hot air balloon hovering 11.1 meters off the ground. Your position is Latitude 42.47 North and Longitude 71.28 West.'
>
> 'You must be an Engineer,' hollered the balloonist from his basket.
>
> 'I am,' she replied, 'How did you know?'
>
> 'Well,' he answered, 'that sounds accurate, but it doesn't help – I'm still lost! I'm sure my friend is upset, you have furthered my tardiness, and now you have also elicited my frustration.'

> The woman below responded, 'You must be in Emotion research.'
>
> 'How did you know?' yelled the balloonist.
>
> 'Well,' said the woman, 'you don't know where you are or where you need to go. You sound concerned about feelings but you aren't actually helping anybody's. And, your primary driving force is a large supply of hot air that appears most effective at propelling you higher."

Psychologists and engineers bring different talents to emotion research. Engineers' inventions allow for huge amounts of objective data to be collected and characterized from people outside labs, out in the field, where emotion happens naturally. Psychologists bring decades of experience understanding human emotion, what elicits it and how to handle the challenges of responding to emotion. What if emotion research could combine the insights and abilities of engineers and psychologists and bring the power of new technology to millions of people, non-psychology students in naturalistic settings, and illuminate individual patterns of emotion, not just group differences? If technology could address the practical problems of handling so many people and their data, how might this change the direction of emotion research?

More collaboration between engineers and psychologists, and more real-world data is what the future of emotion research needs, yet there is more.

Leaving base camp

Emotion is like Everest: Theories and definitions of emotion exist in great abundance, and yet it is still hard to say whether or not a particular thing is emotion, much like climbers at the base of Mount Everest might argue as to whether a particular rock is part of Everest. The fact is you can't define Mt Everest precisely; however, you can still say precise things about it. We can state, without qualification, that Hillary and Norgay climbed Everest in 1953. We can also state precise things about emotion, for example that it significantly influences decision-making and perception, and that an individual's ability to regulate emotion can be dramatically altered by drugs and by lack of sleep. The adventurous researcher can make concrete progress climbing the mountain of emotion research as long as she or he does not get stuck at base camp arguing about whether this or that pebble is part of the mountain.

For a typical emotion researcher, the journey is less like climbing Mt. Everest and more like climbing a very tall ladder, for decades, so it's important to make sure the ladder is placed in a worthy location, so that when you finally reach the end of it, you find the view worthy of your enormous accumulated effort. Where should emotion research ladders be placed? Where is most worthy? Placing the ladder conservatively next to an advisor's or mentor's is the most common method, flattering to them, and this can help launch a career. This approach can explain why some rooftops are overcrowded with people looking down at the same courtyard, polishing and renaming its rocks, quibbling about specs of dirt in colleague's eyes, while some distance away, lies terra incognita, a wild garden of knowledge.

Terra incognita

One highly fertile, vastly underexplored garden is the space of how emotions are elicited, characterized, influenced by and able to influence real-world behavior in environments free from the emotion-shaping influences of "How do I please this experimenter?" and "What are these researchers really testing?" Emotions are about what is real: they change with what *truly matters* to you, which can differ from *what experimenters think matter*s. Consider a participant in our lab engaged deeply in playing a computer game (a task for which he did not have to be paid), maneuvering his character past dangers, around evil enemies, away from gunfire, only to have his character massacred by an explosion. This interaction successfully elicited emotions especially when his character died. However, the biggest peak measured was not during that climax, but was actually much earlier, at a point where, as it happened, the game controls stopped working (Picard, 1997). A broken game controller in the real world can matter more than death in the experiment world.

Collecting spontaneous emotion data from the real world is vital to developing accurate scientific knowledge. Many emotion researchers believe that participants produce "the true smiles of happiness" involving both orbicularis oculi and zygomatic muscles when they are genuinely happy, and in fact people often do smile these Duchenne smiles when genuinely happy, especially in labs when researchers elicit happiness. Repeatedly, however, we have observed genuinely unhappy people making these smiles when they make errors in the real world. This "genuine" smiling more after failure than after success has even been observed in preschoolers (Schneider & Josephs, 1991). Emotion theories need to fit real-world data in general, not just data from lab-elicited emotion.

In one of our earliest studies, a participant dutifully expressed eight emotional states every morning for eight weeks, enabling us to create a machine recognition system that could classify her physiological patterns of emotion, recognizing both valence and arousal, despite daily changes in physiological baselines (Picard, Vyzas, & Healey, 2001). "Yes, I felt anger in your lab when you wanted me to feel anger; I got very good at that" she said; but one day she added, "You know, after I left the lab my fiancé called and I got SO ANGRY! That anger I felt in your lab was NOTHING compared with my boyfriend." When do laboratory engagements really matter to people, and how are the emotions they elicit different?

Changing how much something matters to someone changes the emotion and how it is expressed. Consider the upset driver in the simulator: we see him clench his jaw, take his hands off the wheel, throw them up over his head, swerve off the virtual road, and then flash a Duchenne smile toward the camera or experimenter (and he's not happy.) However, in the real car on the streets of Boston we observed different responses, including negative facial affect and language, tighter gripping of the wheel by both hands and, sometimes, another gesture that it's not necessary to describe in this paper. When we had people drive a circuit around Boston while measuring their physiology, behavior, and driving context, we observed natural expressions that contrasted with those elicited in the lab.

Both situations – simulator and real-world driving – induced emotional distress; however, we now know that conclusions about the real world can be misinformed if based on the simulated.

Figure 1. Left: Traditional electrodermal activity measured from fingertips with electrodes and wires. Middle MIT "Galvactivator" LED reflects wearer's own baseline-corrected skin conductance. Right MIT "iCalm" sensor can be easily worn in daily life to wirelessly gather electrodermal, temperature, and motion data.

For example, we initially expected higher real-world speeds to show higher stress, since this was true in simulators. But straight driving at high speed on the highway elicited low stress compared to straight driving at low speed in the city. Dealing with rotaries and tollbooths was stressful, as expected, but the greatest stress came, surprisingly, when the car was completely stopped in the city, not even at a crosswalk or stoplight. Originally we thought zero velocity was low stress, a time when a driver might perhaps safely take a phone call. However, in the real world peak stress (as measured by autonomic balance and skin conductance) occurred when at zero velocity. Why? Well, it appears to be the case in Boston that people don't like it if you stop for a reason that they cannot see, and this was the case for the driver behind our driver. Furthermore, just as our driver applied her foot to accelerate from zero, a jogger darted out in front of her car. Then when the coast was clear and she started to accelerate, out came a young woman leisurely pushing a stroller, in front of her car. In the real world, anticipation and expectations drive emotions as much as any observable trigger.

Wiring up drivers and cars with a huge array of sensors gave us many new insights into what was eliciting emotions. But we had to wire up the drivers uncomfortably in order to get their physiological state information from the real world. But things have changed since then and now getting data is as easy as slipping on a wristwatch or a Bluetooth ear bud.

A revolution has been happening in sensor and measurement technologies, enabling measurement devices to be deployed comfortably without encumbering daily activity. In heart-rate measurement wearable technologies are available commercially from companies such as Metrisense and Alive Technologies. MIT's mobile electrodermal activity (EDA) and motion sensor (Fletcher et al., To appear.) is now being made by Affectiva. Other companies such as BodyMedia, FitSense, and Polar have developed commercially successful heart rate and activity monitoring products for health and fitness. While none of these measure emotional state directly, these devices do capture physiological changes that co-occur with emotional states, providing objective information related to both arousal and valence, which can be interpreted accurately provided that other influencing variables are held constant (e.g., did the temperature, humidity, electrode pressure stay the same and the EDA go up?) While you cannot perfectly control these variables when participants leave the lab, few controls are perfect in the lab either, and significantly more naturalistic data can be obtained leaving the lab. New technologies allow measurement out in the world, comfortably enough that participants forget they're wearing them: These open up whole new areas of inquiry in emotion research.

Summit: for the people by the people

Emotion research studies rarely benefit their participants. Sure, they pay a modest sum in cash or credits. But consider that many participants have real emotion-related needs; yet, few of them learn anything from participating. While sometimes there are papers published, sometimes participants get a copy, hopefully the paper contains meaningful conclusions, sometimes the results pertain to the participant's group, and sometimes the results help, most times the participants receive no benefit. If they hear their group's results, do they also hear if their personal data was typical for the group, or if they were tossed-out as an outlier or for equipment malfunction? What if the data were inhomogeneous for their group and the results apply to only one cluster, and they do not know which cluster they are in? For many good reasons researchers default to hiding all the individual participants' information. But what if there were a way to protect participants' privacy and let them learn about themselves, without increasing scientific workload beyond what is already overload for most research teams? Could a participant enter a secure ID on the web and see all the anonymized participants' data, and where their data sat relative to the group? Could participants see if they were typical or an outlier, whether the conclusions hold for them, and if any associated recommendations, treatments, etc, might suit them? Social networking tools could also be a part of this process – allowing participants to converse about common features and differences, under pseudonym if desired.

The future proposed here is for better emotion science as well as for better service to people who participate. Consider studies of emotion regulation in autism that have investigated EDA. Many studies have shown that "the autism group has lower autonomic arousal than the typical controls" while others have shown the opposite. Some studies show at least two groups of responders during short lab measurements: one with high arousal and one with low arousal (Schoen, Miller, Brett-Green, & Hepburn, 2008). Meanwhile, using an MIT sensor that measures EDA comfortably all day long, we have found that some autistic people can have extremely high EDA sometimes, and extremely low EDA sometimes, under the same externally induced "baseline" conditions (same person lounging around, appearing relaxed on the outside.) We have also seen cases where a child simply lying down appearing relaxed can have EDA that suddenly escalates by a factor of five or more (Picard, 2009).

Here is a possible explanation for the all-too-common situation where some papers report one physiological finding and some report the opposite: suppose the physiological behavior is dynamically oscillating between high and low, and by using tools that take laboratory snapshots you tend to catch people low or you tend to catch them high, or you tend to catch some of both. In the latter case, you usually don't get publishable results unless you separate your groups (clustering). In the former cases, people publish, and so we see papers where group findings are "too high" or where they are "too low." While we await repeating long-term measurements on many more individuals matched for diagnosis, age, gender, IQ, medications, and so forth, to see if a swinging high-low autonomic pattern is typical in an autism subgroup, if this pattern of both extreme highs-and-lows does not hold for most, does that make it something to toss out? If we find only a cluster of people with such regular swings, wouldn't it be nice to let them know about their individual patterns? Today, there is a tendency to not even consider a finding to be a result unless some gross statistic averaged across a group differs from some other group: researchers have been trained so firmly on constructing tests to see "if $p<0.05$?" that they forget there is enormous value to characterizing dynamic patterns of emotion.

Tools of engineering can do a lot more than discriminate groups: They can characterize complex patterns. Graphical models used to teach computers skills such as speech recognition are increasingly used by engineers for emotion recognition, including facial expressions, prosody, and stereotypical gestures (Albinali, Goodwin, & Intille, 2009; Fernandez, 2004; Kaliouby & Robinson, 2005; Tong, Liao, & Ji, 2007). Now is the time for engineers and psychologists to work together to move beyond simplistic labeling of coarse emotion categories and begin to develop deep characterizations of dynamic naturally occurring emotion.

Speech is nearly universal in people and yet speech varies with individuals and with groups; Emotion may similarly be universal and variable. The research study of individuals is called idiographic while that of averaging across groups is nomothetic. Historically, idiographic researchers published case studies and were accused of bias from their relationship with and subjective impressions of the individual; however, in the last decades the trend has swung to the opposite extreme, where nomothetic researchers ignore individuals and declare no results unless there is a hypothesized group difference confirmed statistically. Is there a way to return to the meaningfulness of idiographic research while preserving the objectivity of nomothetic research?

Increasingly, there is a way, using new technologies that allow ultra dense objective measurements of individuals, coupled with pattern analysis tools that characterize not merely gross statistics like averages, but complex dynamic structures both within and across individuals (Bishop, 2006; Duda, Hart, & Stork, 2001; Jain & Dubes, 1988). Emotion researchers can now learn about individuals using objective long-term measures, and then cluster similar individuals and compare across clusters. Engineering tools of machine learning and multi-dimensional pattern analysis are well suited to characterizing temporal patterns from multiple streams of information, clustering them to identify subtypes, and making cross-cluster comparisons. Moreover, by not averaging away the individual data, each participant's pattern can be preserved, studied, and shared in a way that protects privacy and illuminates relationship to the group. Individual participants can be helped by the findings if they can see their patterns and whether these fit with people who benefited from a particular treatment. Researchers can stop bickering about whether Category A > Category B (unless they like bickering) and instead identify patterns in both categories that are reliable and meaningful for understanding real world needs and behaviors. Both basic science and care for our fellow human beings will benefit with this approach.

Consider the challenges faced by many individuals such as discomfort looking at faces and making eye contact, challenges associated with increased autonomic nervous system activation and hyper-arousal of associated brain regions in autism (Joseph, Ehrman, McNally, & Keehn, 2008; Kleinhans et al., 2009). While almost all of the literature's findings report a group difference almost no articles show that every member of the group has that difference. This phenomenon is common across emotion research: the results hold for the group statistics, but not necessarily for members of the group. If science is to be more than academic, it needs to strive to become increasingly accurate and meaningful for individuals. There was a time when the adjective *academic* meant "of the academy" but now it also means, "having no practical or useful significance" (Merriam-Webster, 2003).

Naysayers to this vision of "emotion research by the people for the people" might argue that the approach further increases the burden for us overworked scientists. More data and more attention to long-term dynamic patterns require more complex time-series and multi-dimensional statistics to learn about, run, compare, and analyze. More real-world data means more noise, labels and context to consider, with more complexity for interpretation. The task can become overwhelming. However, there is another new trend that can help with those last painful steps before the summit of the climb, when energy lags. Here, the power of people online can be harnessed to distribute the load.

Consider the "ESP game" (http://www.gwap.com/gwap/gamesPreview/espgame/) (Von Ahn & Dabbish, 2004) developed to get people to label content in millions of images, a task needed to improve computer vision. Labeling is usually boring, arduous, and error-prone but the ESP game makes it fun to "guess what the other person would label." The results produced orders of magnitude more validated labels than did lab-based approaches. This work has spawned a larger vision of Games with a Purpose (Von Ahn, 2006) that can be used to solve scientific problems by harnessing the (often free!) labor of many people online. Even young children can produce games using online tools such as Scratch (http://scratch/). Our lab has created a free and open source tool for enabling people to label videos online, allowing workload to be distributed worldwide (Eckhardt & Picard). The spirit of making free tools and distributing them online has a great tradition already in emotion research, with experience sampling software available free online (http://www.experience-sampling.org/) (Barrett & Barrett, 2001). Tools forged from collaborating psychologists and engineers can allow more people to participate in research learning about their own emotions, and contributing to accurate and meaningful interpretations of objectively measured data. While the future of emotion research will require more natural data and a lot of work, new technologies are increasingly available to help people efficiently do the work, sometimes even making it fun.

Conclusions

Today when a child teaches a distinguished scientist to upload video on the Internet, when non-researchers can participate in scientific labeling from home, and when gathering autonomic nervous system data 24/7 is as easy as slipping on a sweatband, emotion research is ready for a major leap forward. Ordinary people can gather data, upload it, compare their patterns, share what they learn, and if they wish, share it with scientists for emotion research. Research can be done by the people, for the people. Of course scientists still have to be involved: there is no substitute for deep scholarly study across experiments and for the rigorous development and test of new hypotheses and theories. At the same time, there is no longer any excuse for leaving people out of findings. Emotion research can benefit all its participants, scientists and lay-people, instead of becoming *academic* in the modern definition.

This paper has outlined a vision that begins with partnership between engineers and psychologists – starting at base camp, but not getting stuck there, using technological breakthroughs to move emotion measurement to where people's emotions happen with what really matters to them. The vision includes involving participants in richer ways than as paid subjects, enabling people to learn about their own data and benefit directly from findings in a study, while also contributing to online analysis.

New technologies are allowing regular people to contribute to data submission, annotation, analysis, and more. In this future, individuals can learn how they fit into the larger set of findings, while allowing scientists access to orders of magnitude more natural data than has ever been achievable in lab-based studies. More data, more relevance, more progress, will come to emotion researchers who embrace these new technological capabilities and collaborate fully with the people for whom their research is ultimately intended to serve.

Acknowledgements

The author has never been in a hot air balloon but confesses to making all the mistakes attributed in the story both to engineers and to emotion researchers. Picard also fully discloses that one of the technologies mentioned (iCalm) is to be sold by a company (Affectiva, Inc.) that she co-founded, invested in, and chairs. This work has benefited from discussions and collaborations with Dr. Matthew Goodwin, Dr. Rana el Kaliouby, autistic friends, the MIT Affective Computing group, and colleagues throughout the Media Lab who care about research by the people, for the people. Special gratitude goes to the Nancy Lurie Marks Family Foundation (NLMFF), Microsoft, and the TTT Consortium for helping support this research. The opinions, findings, conclusions and recommendations expressed in this article are those of the author and do not necessarily represent the views of others.

References

Albinali, F., Goodwin, M. S., & Intille, S. S. (2009). Recognizing stereotypical motor movements in the laboratory and classroom: A case study with children on the autism spectrum. *Proceedings of the 11th International Conference on Ubiquitous Computing,* 71-80.

Barrett, L. F., & Barrett, D. J. (2001). An introduction to computerized experience sampling in psychology. *Social Science Computer Review, 19*(2), 175.

Bishop, C. M. (2006). *Pattern recognition and machine learning* Springer New York:.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* Citeseer.

Eckhardt, M., & Picard, R. A more effective way to label affective expressions. Paper presented at the *Affective Computing and Intelligent Interaction,* Amsterdam.

Fernandez, R. (2004). A computational model for the automatic recognition of affect in speech. Ph.D., Massachusetts Institute of Technology.

Fletcher, R., Dobson, K., Goodwin, M., Eydgahi, H., Wilder-Smith, O., Fernholz, D., et al. (To appear.). iCalm: Wearable sensor and network architecture for wirelessly communicating and logging autonomic activity. *IEEE Transactions on Information Technology in BioMedicine,*

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data.*

Joseph, R. M., Ehrman, K., McNally, R., & Keehn, B. (2008). Affective response to eye contact and face recognition ability in children with ASD. *Journal of the International Neuropsychological Society, 14*, 947-955.

el Kaliouby, R. , & Robinson, P. (2005). Real-time inference of complex mental states from facial expressions and head gestures. *Real-time vision for human-computer interaction* (pp. 181—200) Springer-Verlag.

Kleinhans, N. M., Johnson, L. C., Richards, T., Mahurin, R., Greenson, J., Dawson, G., et al. (2009). Reduced neural habituation in the amygdala and social impairments in autism spectrum disorders. *AJP in Advance, 166*(4), 467-475.

Merriam-Webster, I. (2003). *Merriam-webster's collegiate dictionary* Merriam-Webster.

Picard, R. (1997). *Affective computing*. Cambridge, MA: MIT Press.

Picard, R., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions of Pattern Analysis and Machine Intelligence, 23*(10), 1175—1191.

Picard, R. W. (2009). Future affective technology for autism and emotion communication. *Philosophical Transactions of the Royal Society B: Biological Sciences, Dec 2009.*

Schneider, K., & Josephs, I. (1991). The expressive and communicative functions of preschool children's smiles in an achievement situation. *J. Nonverbal Behavior, 15*, 185—98-185—98.

Schoen, S. A., Miller, L. J., Brett-Green, B., & Hepburn, S. L. (2008). Psychophysiology of children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 2*(3), 417-429.

Tong, Y., Liao, W., & Ji, Q. (2007). Facial action unit recognition by exploiting their dynamic and semantic relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(10), 1683.

Von Ahn, L. (2006). Games with a purpose. *Computer, 39*(6), 92-94.

Von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 319-326.