Understanding Digitally-Mediated Empathy: An Exploration of Visual, Narrative, and Biosensory Informational Cues

Max T. Curran
BioSENSE, School of Information,
University of California, Berkeley
Berkeley, California, USA
mtcurran@ischool.berkeley.edu

Jeremy Raboff Gordon
BioSENSE, School of Information,
University of California, Berkeley
Berkeley, California, USA
jrgordon@berkeley.edu

Lily Lin
School of Information, University of California, Berkeley
Berkeley, California, USA
lilyelin@ischool.berkeley.edu

Priyashri Kamlesh Sridhar
Augmented Human Lab, Singapore
University of Technology and Design
Singapore, Singapore
priya@ahlab.org

John Chuang
BioSENSE, School of Information,
University of California, Berkeley
Berkeley, California, USA
chuang@ischool.berkeley.edu

ABSTRACT

Digitally sharing our experiences engages a process of empathy shaped by available informational cues. Biosensory data is one informative cue, but the relationship to empathy is underexplored. In this study, we investigate this process by showing a video of a "target" person's visual perspective watching a virtual reality film to sixty "observers". We vary information available to observers via three experimental conditions: a baseline unmodified video, video with narrative text, or with a graph of electrodermal activity (EDA) of the target. Compared to baseline, narrative text increased empathic accuracy (EA) while EDA had an opposite, negative effect. Qualitatively, observers described their empathic processes as using their own feelings supplemented with the information presented depending on the interpretability of that information. Both narration and EDA prompted observers to reconsider assumptions about another's experience. Our findings lead to a discussion of digitally-mediated empathy with implications for associated research and product development.

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300844

CCS CONCEPTS

• **Human-centered computing** → **Social content sharing**; **Empirical studies in collaborative and social computing**; *Laboratory experiments*; *Empirical studies in HCI*;

KEYWORDS

social computing, computer-mediated communication, biosensing, empathy

ACM Reference Format:
Max T. Curran, Jeremy Raboff Gordon, Lily Lin, Priyashri Kamlesh Sridhar, and John Chuang. 2019. Understanding Digitally-Mediated Empathy: An Exploration of Visual, Narrative, and Biosensory Informational Cues. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300844

1 INTRODUCTION

Research Motivation and Goals

Scientific meta-analyses of historical trends suggest that levels of empathy in the United States may be at a low [20]. Ideological and political divides are growing and a critical eye has settled upon social networking services for their potential role in reinforcing this rift as a byproduct of increasingly filtered and curated social environments functioning as echo-chambers [3, 8]. More clarity and flexibility communicating one's inner thoughts and context enables improved understanding by others. The most direct version of this process occurs during in-person interpersonal communication in which one person describes their experience to another using a rich set of verbal and nonverbal cues such as facial expression, physical gestures, touch, voice, words, posture, etc. We can relate to one another by understanding each others' perspectives and reactions to experiences via these cues which are shaped by our identities, histories, and relationships.

Digital communication frees us of the requirement to be physically collocated, but comes at the cost of limiting what cues are available to convey an experience. The availability of these cues has been shown to have effects on crucial social processes like trust development [5]. Users of social media and video sharing and streaming platforms like Facebook, Instagram, Youtube, and Twitch participate in mediated experience sharing. In many cases, the images and videos posted to these sites are augmented with additional information like text descriptions, emojis, streaming video of their faces, and more. Biosensory data is one such information type recently explored in HCI research which may afford a unique kind of additional context–one that may bear directly on internal experience.

Building upon research from HCI and psychology in the areas of empathy and mediated interactions, we conducted an experimental study to investigate how different informational cues about a target individual's experience influence observers' accuracy predicting the target's affect. Additionally, we interviewed observers regarding their impressions and use of these cues when making their predictions about the target. Because previous research has indicated that target-observer dyad characteristics such as race [31], motivational differences by gender [17], class [21], and relationship closeness [9] affect empathic processes and could reinforce stereotyping by the observer, the present study focuses on anonymized communication in order to better isolate the effects of different informational cues as well as explore anonymity as it is uniquely possible in mediated communication. Compared to other informational cues for empathy like facial expression and voice, biosensory information in the way it is employed here does not clearly disclose one's identity.

Our aim with this research is to offer insights into the fundamental social dynamics of an increasingly technologically-mediated world. What qualities affect how well someone understands our emotional states in digital communications? Are some types of information more apt than others for conveying our emotions and engendering an accurate interpretation by another? Could surfacing and sharing typically invisible information about our bodies' emotional responses augment this process? This study addresses these important questions with implications for the design and further research of communication systems for fostering empathy.

Related Work

Motivated Empathy & Empathic Accuracy

The ability of individuals to empathize in various scenarios is a complex but long studied topic. Our modern understanding of empathy involves two key components: observers's *perception and understanding* of a target's affective state, and observers *feeling* targets' affective states, referred to as cognitive and affective empathy, respectively. Furthermore, recent research has questioned the classical assumption that empathy is an immutable trait. Instead, "motivated empathy" proposes that the process of empathy is at least partly under our control and further delineates subcomponents related to cognitive empathy: "mind perception" the detection of another's internal states, and "mentalizing" the drawing of explicit inferences about another's emotions [32]. Part of this theory includes possible strategies to control empathy including "attention modulation", that observers can increase or decrease empathy by adjusting their attention to social targets' emotions; and "appraisal" in which "observers might shift their beliefs about the intensity of targets affective states". These ideas echo findings from social neuroscience research showing the brain's pain matrix activates when one person is cued that a close other is experiencing pain [7, 28] and that activation is modulated by the observer's perceived intensity and saliency of the pain [2]. These motivational qualities of empathy provide an opportunity for designing technological systems that shift the attention of an observer toward affective cues and augment their abilities of mind perception and mentalizing.

Our abilities to meaningfully and productively engage with one another for social understanding and support, whether we are distressed or filled with joy, are predicated upon how well we can both convey our own internal states as well as understand the internal states of others. This skill in "everyday mind reading" is referred to in psychological research as "empathic accuracy" (EA) [18]. By measuring the effects different informational cues have on EA, we can better design for the cognitive component of empathy in mediated communication.

HCI & Social Biosensing

In the field of HCI, compelling designs of technological system and applications explore new platforms and channels for empathic communication. Social perspective taking, defined as "the process of an 'observer' discerning the thoughts, feelings, and motivations of a 'target' " [10], has been recently explored using virtual reality (VR). An exemplary case is one series of experiments indicating improved helping behavior toward individuals with colorblindness after an embodied simulation of being colorblind using VR [1]. Recent directions in affective computing encourage an interactional, rather than informational, model of emotion representation in HCI systems in which users are actively engaged in interpreting of emotion-related data alongside situational context, as opposed to relying on hardline system-generated classifications [4]. Intentional use of ambiguity in the design of HCI systems has been proposed as a way to enable and encourage this active consideration and meaning-making by users [12], a process particularly relevant to an arena as multifaceted as the interpretation of others' emotions.

Slovák et al. showed evidence linking consistent electrodermal activity (EDA), small changes in the moisture of the skin related to emotional arousal, synchronicity between pairs of individuals conversing in real-world settings with high emotional engagement and described applications for social skills learning and enhancing remote communication [30]. Some of the same researchers explored the sharing of heart rate data to improve social connectedness and reduce loneliness, finding distinctions between the perception of this data as information versus connection [29] and suggesting heart rate as a useful cue of intimacy that can function similarly to gaze and interpersonal distance when attributed to a target [19]. Further still, exploratory design work has been done with wearable devices allowing interpersonal sharing of breathing motions [25] and EDA-modulated color changing fabric to study wearers' and observers' interpretations of this ambiguous signal [16]. One study testing social biosensing in the wild built an app that allowed users to share their heart rate with an accompanying message, finding that context was key to using this biosignal for emotional disclosure [23]. The importance of context comes up again in an investigation of the expressivity of electroencephalography (EEG) data, which also explores a range of visualizations from a raw graph to a set of emojis finding the majority of participants preferring the subjective interpretability and information density of the more raw forms [22]. Another in the wild implementation of heart rate sharing in text communication found it allowed conversation partners to better perceive when the other was angry or excited and as well as a desire among users to see and interpret even small fluctuations in the signal and to see their partner's real-time reaction to reading their message [15]. The present study is intended to compliment the largely qualitative findings of these works with a quantitative investigation of social biosensing effects on empathy. An intriguing duality of biosignals is their status as both intimate and analytical; while they're sourced from the innermost sensations of a person's body, the forms in which they are often represented or explained like numbers and graphs give them medical or scientific quality. Therefore in relation to empathy, we expected biosignals represented in this way to affect cognitive (rather than affective) empathy, which informed the focus of our study.

In a review including many of these examples, Chanel et al. emphasize the potential of displaying normally hidden cues like physiology for social interaction and connection [6]. In their study of perspective taking and virtual reality, Gehlbach et al. describe evidence suggesting perspective taking can, counterintuitively, reinforce stereotyping if the target is highly stereotype-consistent [13]. Given the possibilities of anonymity in computer-mediated interactions, an exploration of perspective taking via more anonymized forms could be an avenue to reap the benefits while avoiding this drawback. This emerging field of research is promising, but the dynamics and efficacy of sharing experiences augmented by biosensory information have yet to be examined through the lens of empathy which is the aim of the present research.

Research Questions & Hypotheses

Following the notion that empathy has a motivational component as well as studies from HCI demonstrating the empathic potentials of sharing biosensory data and taking another's perspective in VR, we carried out an experimental study paired with semi-structured interviews to answer our research question:

RQ: How does the type of information presented about a target's experience: (field of view alone ("FoV"), the addition of a textual narrative ("Narration") or electrodermal activity ("EDA"), affect (1) the correlation with and (2) the difference between target's self-ratings and observers' estimations, as well as (3) their feelings of empathy toward the target?

We expected Narration to achieve the highest accuracy, lowest error, and highest feelings of empathy given the target's high degree of control over this cue as well as the relative ease of interpreting written text. EDA we expected to perform in the middle as it provides indirect information about the target's internal state and as demonstrated by previous research can increase connectedness and intimacy, though it is less controllable and straightforward compared to narration. With no additional information about the target's experience present, we expected FoV to result in the lowest accuracy, highest error, and lowest feelings of empathy. Thus our hypotheses for the quantitative explorations can be summarized as:

H1: Narration > EDA > FoV for observer EA
H2: FoV > EDA > Narration for observer rating error
H3: Narration > EDA > FoV for observer feelings of empathy

Additionally, we sought to gain qualitative understanding of observers' impressions, perceptions, and usage of different information types through semi-structured interviews with observers in all conditions. While the quantitative measures allowed us to address our specific hypotheses, they are not well-suited to capture the many facets and nuanced nature of empathy which the interviews could allow for.

2 METHODS

Participants and Data Collection

Study participants were recruited via e-mail through an on-campus experimental lab at a large public university. 60 participants took the role of "observers" attempting to assess the feelings of a target using a stimulus determined by each of the three study conditions. Three other participants were removed after researchers agreed they did not sufficiently understand task instructions. The only selection criteria beyond those who responded to the e-mail advertisement was the ability to speak and understand English and not having seen the film used in the study before. Observers were randomly placed into one of the three study conditions. These 60 participants had a mean age of 20.7 ± a standard deviation of 3.1, 45 identified themselves as female, and 13 as male. All participants were undergraduate or graduate students at the university where the study was conducted.

From a separate set of four pilot participants, one was selected to act as the observers' "target". This target provided data about their experience which was used to create stimuli and compare with observers' assessments to calculate outcomes like EA. This target was not intended to be representative of all possible users, but instead functioned as an example to generate stimuli that would allow comparisons between observers receiving different information types, the main manipulation of this study. This participant scored moderately on emotional expressivity, was at least moderately expressive in their narration about the VR experience, and had a suitable amount of variation in their EDA data recorded during the VR experience, and was therefore deemed suitable for selection.

After agreeing to an electronic informed consent form via the recruitment e-mail, participants completed a set of online questionnaires which included demographics and the Berkeley Expressivity Questionnaire (BEQ) [14]. Participants were paid $5 for completing these online questionnaires, and were subsequently invited to participate in the in-person portion of the study for which they were paid an additional $20 for completing.

The in-person portion of the study was split into two phases: one to gather information from the target participant to generate stimuli, and a second to expose this stimuli to observers as they complete an empathic accuracy task followed by a semi-structured interview. Timelines of these two phases are shown in Figure 1. The recruitment process and study protocol was approved by the local ethics review board.

Target Phase. In the first phase the target experienced a short immersive VR film presented using an HTC Vive while wearing an Empatica E4 wristband which recorded their EDA. The participant was seated in a swivel chair and was instructed they could look around during the film but not to move from the chair. The use of VR allowed a greater level of immersion and presence for the target, and enabled seamless recording of the target's visual perspective to show to observers as a simulation of the online practice of sharing photos or video of one's first-person experiences. For this study the film *Pearl* was selected, an award-winning 6 minute immersive animated film released as part of Google's Spotlight Stories series. The film is available for download free of charge on the HTC Vive store and a 360 degree YouTube video version is also available [1]. The film tells the story of a father and daughter through snapshots of the daughter's childhood and teenage years while the two lived out of their car. This film was selected for several reasons: it allows for exploration in a seated position without too much movement to alleviate motion effects on biosensory data capture, the various scenes in the story were likely to elicit a range of positive and negative emotional responses, and the family-focused theme was thought to be widely relatable to a young adult audience such as the participant pool.

An audiovisual recording of the target's field of view was produced and subsequently played back for the target on a 2D laptop computer screen. While watching the playback, the target was asked to narrate how they were feeling throughout the experience. The audio of this narration was also recorded. The target then performed a cued-recall task, further described in [24], in which they used a USB rating dial to continuously rate the valence of their experience on an 11-point on-screen scale labeled Very Negative (-5), Neutral (0), and Very Positive (5) as they watched the video a second time. Prior to this activity they completed a short 30-second practice and were given a chance to ask questions or clarify the task. The video presentation, narration recording, and dial ratings were collected using PsychoPy [26]. Finally, the target provided ratings on how they felt overall for 17 discrete emotions: aesthetic appreciation, amusement, anger, anxiety, awe, boredom, calmness, confusion, disgust, excitement, fear, interest, joy, and nostalgia, each assessed on 7-point Likert-type scales.

Observer Phase. The primary manipulation for this study was the type of anonymous information observers were given while assessing the target's original experience. These information types constituted the three conditions of the study:

[1] https://www.youtube.com/watch?v=WqCH4DNQBUA

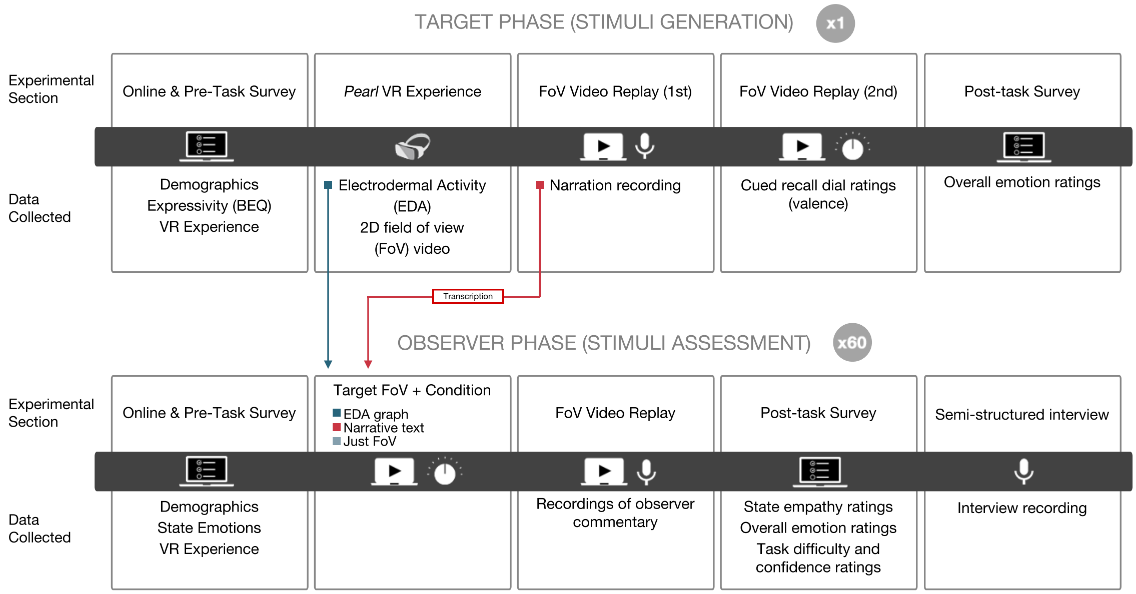


Figure 1: Timeline of study procedures for the two phases of the study: the target phase where the target experienced a short film in VR and data was gathered about their experience to generate stimuli, and the observer phase where empathic accuracy was assessed by study condition with a rating task and follow-up questionnaires and semi-structured interviews.

(1) visual perspective via a recording of the target's VR headset field of view ("FoV"), (2) FoV accompanied by the target's descriptions of their experience in the form of on-screen video subtitles ("Narration"), or (3) FoV accompanied by an animated graph of the target's EDA data ("EDA"). Example screen shots from each of the three conditions are shown in Figure 2. The stimulus for the Narration condition was produced by adding subtitles of the target's narration using the Aegisub software. Subtitles were timed generally to the speed which the target dictated, with some slight adjustments to ensure they were on the screen long enough for observers to read. The stimulus for the EDA condition was created using the Matplotlib Python package to plot the target's raw EDA during the film and overlaying this animation onto frames of the video using the MoviePy Python package. An empty plot area is shown at the start of the video with the Y axis labeled "EDA" and values of 0, 1, and 2 and a line showing the target's content-synchronized EDA populates in real-time as the FoV video plays. The EDA condition was included to build upon existing research in social biosensing using a data type that is highly related to emotional experience and ambiguous enough to allow for engagement and interpretation by the observers. The Narration condition simulates an already common informational cue in mediated communication, such as the inclusion of descriptive text or caption for a photo or video, but in an anonymous and continuous format to be comparable to other conditions. Finally, the FoV condition was included as a control with no additional information aside from the visual perspective of the target.

Observers were asked to do a task similar to the cued-recall dial ratings the target completed, but instead of rating their own feelings, they were asked to perform an empathic accuracy task modeled after [33], in which they rate how they thought the original viewer was feeling on the same 11-point scale from Very Negative to Very Positive continuously throughout the video. A photo of the setup for the rating dial tasks is shown in Figure 3. While this task is somewhat complex, requiring the observer to watch the video for the first time and make their ratings, this situation is most similar to mediated communication scenarios where a recipient simultaneously parses their own and the sender's experience.

Like the target, observers did a short 30-second practice with an unrelated video before the task and were given the opportunity to ask questions or clarify the task. Just before beginning the empathic accuracy task, participants read a short set of instructions that varied by condition: for the FoV condition this was a simple reminder to rate how they think the original viewer was feeling. For the Narration condition this reminder included a note that subtitles would be shown along the bottom of the screen provided by the original viewer about how they were feeling. Finally, for the EDA condition the same reminder was included as in the FoV condition, along with the following description:

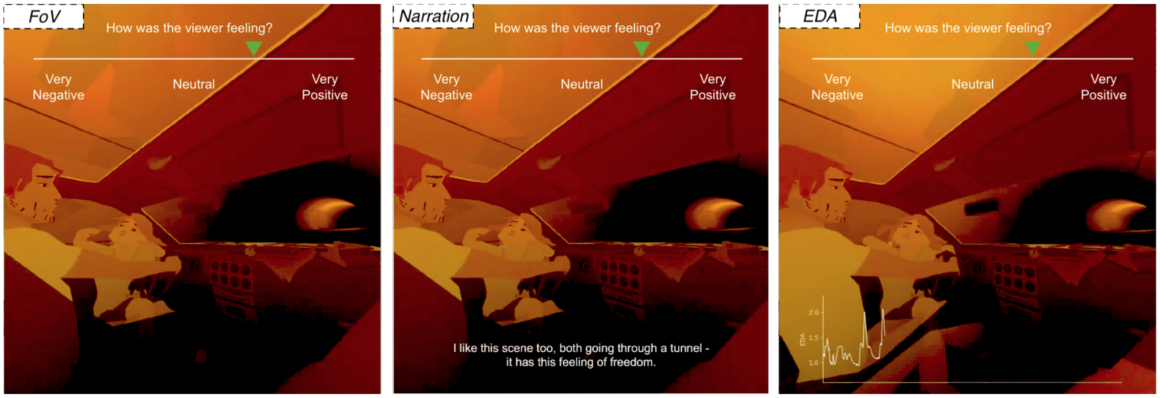


Figure 2: Example screen shots from the stimuli shown to participants during the rating task for each of the conditions: Field of View (FoV), Narration, and EDA. All conditions displayed the rating scale along the top of the screen, which moved when the participant adjusted the USB dial. The Narration condition showed text transcribed from the target's narration of their experience, much like subtitles. The EDA condition showed an animated graph of the target's EDA which was populated as the video progressed, displaying the current and historical EDA data of the target.

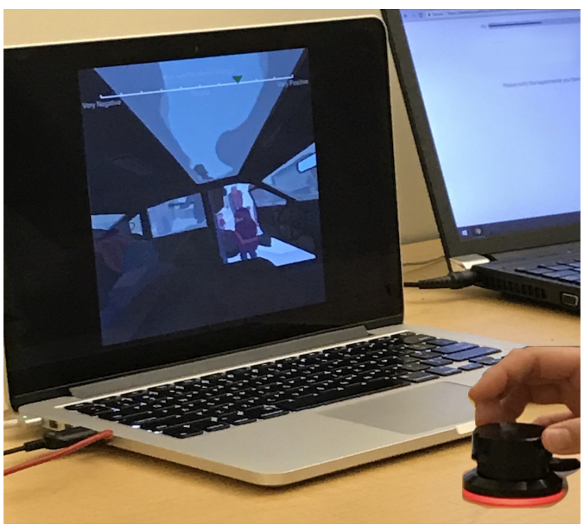


Figure 3: Experimental setup for both the target's self cued-recall rating task and the observers' empathic accuracy task. The USB rating dial can be seen in the bottom right, which controlled the position of the selector on the scale titled either "How were you feeling?" (for the target) or "How was the viewer feeling?" (for observers), ends labeled "Very Negative" and "Very Positive".

> *In the real task you will also see a moving graph of the viewer's electrodermal activity (EDA) while they were in the VR experience.*
>
> *EDA is a measure of changes in moisture or sweatiness of the skin. The meaning of EDA is not fully understood, but increases in EDA have been associated with different forms of arousal such as fear, surprise, stress, or excitement.*

After the rating task, observers answered a set of questions including two "state empathy" questions from [11] to assess feelings of empathy: "I felt as though I were in the viewer's shoes" and "I imagined myself in the viewer's situation". The post-task questions also included overall ratings of the target in the same 17 7-point Likert-type scales the target completed, and two additional 7-point Likert-type scales about the difficulty of the rating task and their level of confidence in their ratings. Finally, each observer participated in a semi-structured interview probing general impressions, their strategy in the rating task, and what they thought of the EDA and narrative information. Interview questions are listed in Appendix A.

Data Analysis

Empathic accuracy (EA) was calculated as the Pearson R coefficient of the time-series correlation between the target's cued-recall dial ratings and an observer's mirrored time-series of ratings of how they thought the target felt. In addition, a measure of observer error was calculated as the difference between observers' and target's ratings. As a correlation, values in the EA measure are more sensitive to directionality of rating changes, whereas the error measure is more sensitive to absolute differences between the two time series.

We calculated means across participants in each condition for these values and conducted ANOVAs as omnibus tests of statistical difference, which, if below the chosen significance threshold of $p < 0.05$, were followed up with pairwise t-tests.

The same difference of means testing was also carried out for the post-questionnaire items of state empathy, overall emotion rating accuracy, task difficulty, and reported confidence in ratings. The audio recordings for the semi-structured interviews were transcribed and qualitatively coded by two of the researchers resulting in the themes presented in the results.

3 RESULTS

Empathic Accuracy

Among participants in the FoV condition, mean EA was 0.534 $\pm$ a standard deviation of 0.204, with Narration achieving a mean EA of 0.663 $\pm$ 0.107, and EDA with 0.427 $\pm$ 0.196. A one-way ANOVA among these conditions resulted in significance ($F = 8.117$, $p < 0.001$), and follow-up pairwise t-test comparisons were also all significant: FoV and Narration ($T = 3.439$, $p < 0.001$), FoV and EDA ($T = 7.021$, $p \ll 0.001$), and EDA and Narration ($T = 6.097$, $p = \ll 0.001$). Figure 4 shows mean observer ratings by condition over the course of the film duration as well as the target's self-ratings, paired with annotations detailing events in the stimuli. The left panel of Figure 5 shows the significant difference in overall means between conditions. These results reject part of H1 which predicted EA would be higher in the EDA condition than in the FoV condition.

Rating Error

Among participants in the field of view condition, mean observer error was 1.759 $\pm$ 0.570, for participants in Narration 1.553 $\pm$ 0.419, and EDA with 1.759 $\pm$ 0.556. A one-way ANOVA among these conditions was not significant ($F = 0.892$, $p = 0.416$). The right panel of Figure 5 shows mean rating error by condition. These results reject H2 which predicted FoV to have the lowest error, followed by Narration, and finally EDA.

Post-Rating Task Measures

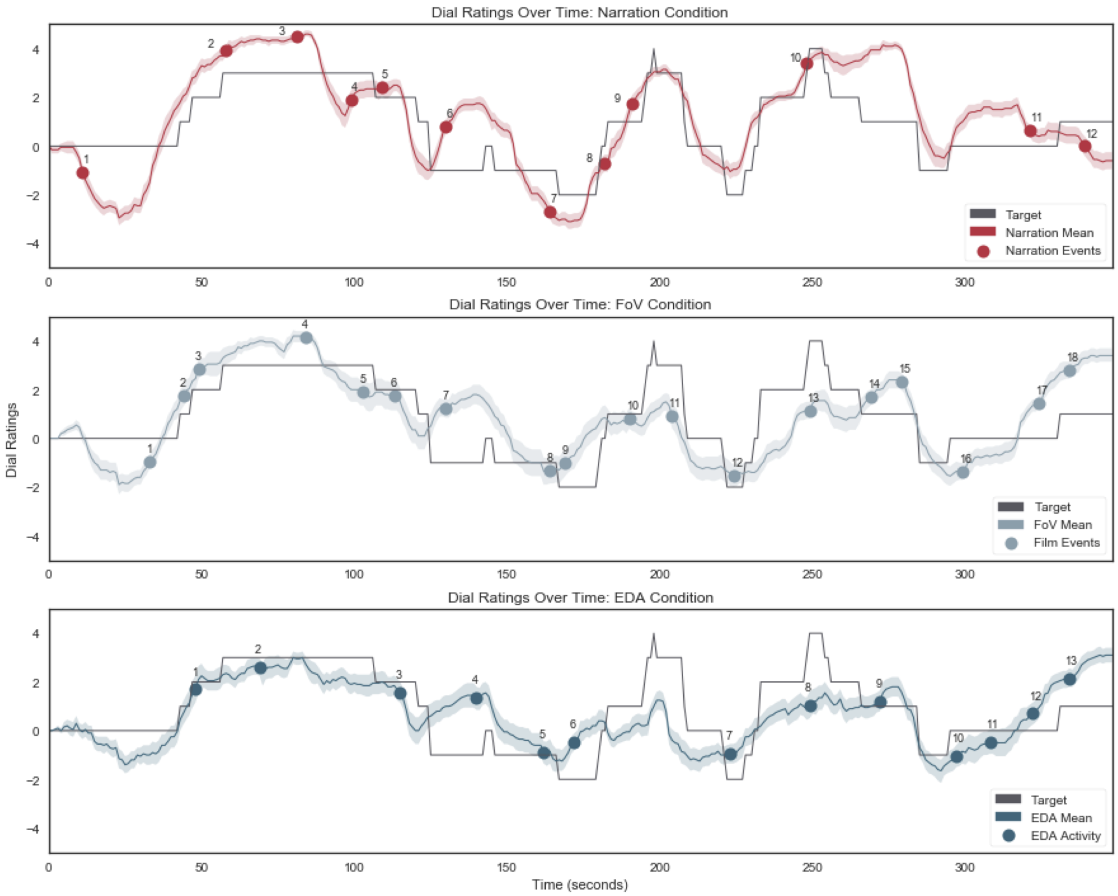
Figure 6 shows mean values for four post-rating task measures observers completed: the mean of the two state empathy questions, mean accuracy labeling overall emotions of the target, rating task difficulty, and observers' reported confidence in their ratings. None of these measures resulted in significant ANOVA tests, indicating no statistical differences between conditions for these values which includes the rejection of H3 regarding state empathy differences between conditions.

Qualitative Findings

Interview transcripts were analyzed to identify emerging themes on strategies employed by participants to gauge the target's emotions, their opinions on the EDA and narration information, as well as general impressions of the tasks.

Acute Attention to Available Cues. In all conditions participants reported reading into various cues that were available to them to rate the target's feelings. In addition to this, they also relied on cues from the elements of the content itself such as color, lighting, music and facial expressions of the characters in the story, to infer about the target's feelings. In response to what strategies they used when predicting the target's feelings one participant stated: "I thought it was a little difficult to interpret but with the color schemes and all the facial expressions, I was able to get what was going on a little bit". Participants were very attentive to specific features of the supplementary information whether it was narrative information or the EDA sensor data. Such cues also seemed to serve as a reminder to rate the target's feelings rather than their own. In the Narration condition, participants isolated adjectives as important cues for how the viewer was feeling: "I paid a lot of attention to the adjectives or the way that they said they were feeling and then I used that to guide the rating" or, "... like every time (they) say there's surprise or just little words like that, I thought the text was super helpful in helping with the flow of the video, sometimes when I was unsure what to feel I just referred to the text and it cleared things up". Participants who saw the viewer's EDA data as supplementary information were attentive to peaks in the graph, though reported that it wasn't always helpful in their task: "when it spikes up or down that's when I would really pay attention to see why the viewer felt like that, had that spike". It's clear from the significantly lower EA in the EDA condition that even when many participants said they generally ignored the EDA graph, the presence of this data did affect their ratings in some way.

Mind Perception via Narration and EDA. Another similarity between the conditions was their function to remind the observer of the existence of the original viewer, that they may have different reactions than themselves and forced them to question their assumptions about the viewer's experience or identity. One participant in the Narration condition reflected on the power of the narration to guide their ratings beyond their own reactions to the film: "I couldn't let my feelings get in the way and had to look at it from what I was reading about how the narrator felt". A participant in the EDA found they were affected by the EDA graph in a similar way, noting that the activity of the graph where they didn't expect it led them to consider how the target was different from themself: "sometimes I felt like there were some spikes where I thought there shouldn't be but maybe it was because the viewer was a parent and maybe they would know some things that kids don't... Maybe that's why". In both of these cases and many others, the additional information prevented observers from



Film Events (all conditions)		
1. Flashback starts **2.** Father begins playing his song while driving with young daughter in the backseat **3.** Young daughter catches fireflies through moon roof **4.** First scene of busking **5.** Busking outside in the snow **6.** Father sees another family in their home and dons a sad expression **7.** Young daughter plays with other children in the car	**8.** Teenage daughter ignores father in the car, changes car radio abruptly **9.** Teenage daughter plays guitar and sings her father's song alone in the car **10.** Teenage daughter shoots fireworks off in the parking lot of a gas station with friends **11.** Teenage daughter is pulled over while driving and then argues with father, storming out of the car **12.** Father finds teenage daughter after searching for her, dons expression of relief	**13.** Teenage daughter and her friends park the car on a beach and go swimming **14.** Teenage daughter drives through tunnel as she did with her father when she was young **15.** Car breaks down and gets towed **16.** Flashback ends **17.** Adult daughter drives the car to show her father **18.** Adult daughter with her father and friends surrounded by camera flashes and cheers
Narration Text (narration condition only)		**EDA Activity (EDA condition only)**
1. "At this point I'm like "What's happening?". It seems like I'm in the front seat of a car, looks quite tattered - did something go wrong? I don't know." **2.** "I really like this scene when she kind of catches the light. That's probably one of my favorite scenes actually, the catching light thing." **3.** "They seem to have a really close relationship that makes me feel happy." **4.** "But still they have a pretty good relationship and that makes me feel good." **5.** "They're busking together and they've busked before so it makes me feel happy to see both of them doing it together even though they seem not to have a permanent place." **6.** "Here it makes me feel a little bit sad, but then I see she's got other friends coming in so I was like yay new characters to this film - cool."	**7.** "Now she enters the car all by herself, it's pretty sad. Where's dad? Oh no…" **8.** "I get to see all her friends - that guy is really funny so it makes me feel good." **9.** "Friends hanging out, playing music together. It's really nice seeing the fireworks go off - that made me feel excited." **10.** "I like this scene, this is another one of my favorite scenes with all of them dancing outside of the car." **11.** "I didn't quite understand this part - the girl here looks quite different... and yeah I dunno." **12.** "I feel like this was supposed to be a happy scene but I couldn't really relate to it, didn't quite understand why people were standing around the car and cameras were flashing."	**1.** Major EDA spike **2.** Major EDA spike **3.** Major EDA spike **4.** Major EDA spike **5.** Major EDA Spike **6.** EDA drop **7.** Minor EDA spike **8.** Minor EDA spike **9.** Minor EDA spike **10.** Minor EDA spike **11.** Minor EDA spike **12.** EDA drop **13.** EDA drop

Figure 4: The target's self-ratings during *Pearl* alongside mean observer ratings in each condition, with standard deviations shaded. Significant film events, present in all three conditions, are annotated in the FoV center panel. Narration text and EDA activity events are annotated in the Narration (top) and EDA (bottom) panels. Annotations are detailed in the lower table.

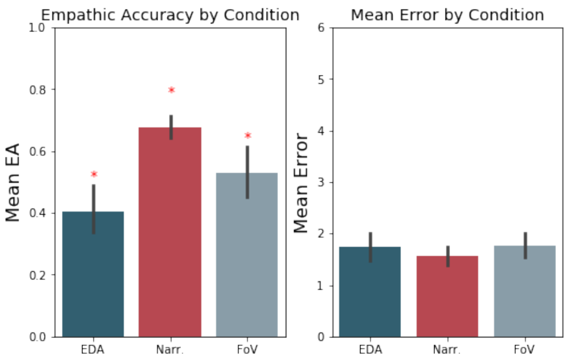


Figure 5: (Left) Mean EA by condition, ANOVA significant (p<.001) as well as pairwise t-tests: Narr. > EDA (p<.001), Narr. > FoV (p=0.005), FoV > EDA (p=0.039). (Right) Mean error by condition, ANOVA not significant (p = 0.416).

relying solely on their own reactions to the film content contrary to participants in the FoV condition who had no supplementary information.

Authority Attribution by Cue Interpretability. While the conditions shared similarities, a particularly interesting distinction between the narrative and EDA information types was their respective utility and the level of authority observers attribeud to them when describing their rating strategies. In the EDA condition, observers spoke about referencing the graph secondarily to the content of the video, allowing the context to be the primary dictator of the viewer's experience and the EDA graph to offer auxiliary input of lesser importance to the rating task, as one participant in the EDA condition described: "whenever I saw something that would [make me] change the position of the dial, I looked at the graph to see if it changed, the viewer's EDA. If it did, and I agreed, then I would move it in the direction that I thought that the spike indicated". Here the participant first watched for changes in the film that they felt would cause an emotional response and only then would look to the EDA graph activity for confirmation. In the Narration condition, however, the relationship between the context and supplementary information was swapped, where a common strategy involved a much higher reliance on the narrative text and secondarily on the context to confirm their interpretation of the viewer's words as a participant in the Narration condition stated: "when she said that it feels pretty sad, I was convinced that she was feeling negative and I moved it further... So I used a little bit of what I could see in the visual like the expressions but especially more of what the text was saying". In contrast to the EDA condition, this participant's response indicates they predominantly used the narrative text to guide their ratings, supplemented by their interpretation of the film content. It's important to note the narrative nature of this film did allow many participants to interpret an intended emotional response without any supplementary information specific to the target. Several participants across conditions described rating what an "average person" might feel, as a functional proxy for the target's experience.

Reactions to an Unfamiliar Signal. Participants had notable reactions and interpretations of the EDA information in particular, likely because it is a novel type of data not commonly encountered outside of research. Participants generally had little to no prior experience with EDA, and thus based their knowledge of the signal solely on the short, purposefully non-specific instructions provided to them before the task. Some participants chose to ignore the data given the inexact description while others were thoughtfully skeptical as described by one EDA participant: "there are people who will get sweaty y'know because they're excited and they're nervous so you can mix those emotions so I feel like the graph isn't too specific with how they're feeling at the moment just kind of gives you this graph". Here the participant discerned that the EDA information did not provide a straightforward indicator of the target's feelings given their own knowledge of mixed emotions. Some interpreted the signal based on their own experiences with sweat, as one participant describes: "I thought that the higher your EDA the more moisture in your skin you might produce was usually like when you're worrying, when you're anxious, you get sweaty clammy hands that sort of thing". This participant's association of sweat with anxiety or nerves clearly affected their interpretation more than the description of EDA provided that stated EDA changes can also be due to positive experiences like excitement. For many participants the target's EDA graph did not match their expectations given their own interpretation of the content, as one describes their process: "I was noticing where it peaked, and I didn't really know why it peaked at this particular scenes. I still don't know why. Because I didn't feel any specific emotion during those specific scenes. I felt some of the more emotional scenes were elsewhere in the video. I didn't see the relationship". Participants like this one talked about placing some trust in the EDA data to aid them in their ratings at the beginning of the task, but due to the signal's ambiguous and confusing nature they were easily swayed to discount its veracity when they saw unexpected EDA activity where they didn't expect any given their own interpretation of the content.

Observer Cue Suggestions. One interview question asked participants to think of other types of information about the target that would have helped them rate their feelings. Some participants reported they would want more information about who the target is to aid in understanding their reaction. Other participants described various data streams

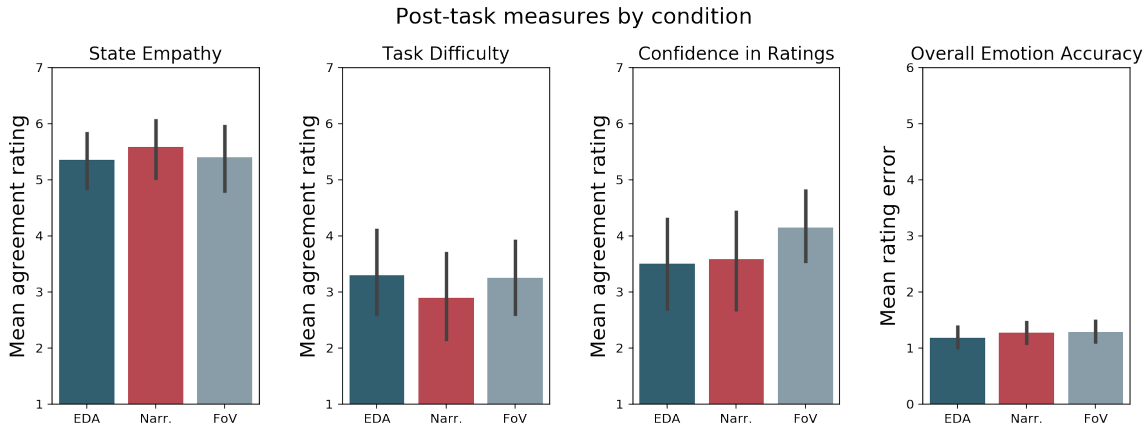


Figure 6: Results of four post-rating task measures by condition. From left to right: mean responses to the two state empathy questions, mean difficulty ratings, mean confidence ratings, and mean rating error when labeling 17 discrete emotions overall. ANOVAs across conditions indicate no significant differences between conditions among these four measures.

including facial expression, speech and other sounds like breathing or gasps, and one participant mentioned heart rate. Here a participant describes using audio information for similar purposes: “maybe audio as well, like if they expelled a lot of breath, they’d be pretty shocked, or I dunno if someone’s really emotive they might say “Aww” or something like that [...] like when you’re in a movie theater and you hear people make sounds, you can understand their reactions”. Vocal exertions like those described by this participant may also allow for the advantages of anonymity and automatic recording and embedding and are worth future exploration as a type of informational cue.

4 DISCUSSION

Attention Modulation in Mediated Communication

In all conditions, participants reported keen attention to details of the content including lighting, color, music, and character expressions and of the supplementary information like particular adjectives in the narration or small fluctuations in the EDA graph. Participants’ reported engagement with these various cues hearkens to the attention modulation strategy Zaki proposes as a way to up or down-regulate empathy [32]. In face-to-face scenarios using this strategy to increase empathy would involve increased attention to a person’s appearance or behavior, and in our anonymous setting where these cues are unavailable this appears to be mirrored instead using cues from the content, emotion-laden words or phrases narration, or EDA graph activity. This is a reminder of the importance of carefully considering the informational cues afforded by a system and to not underestimate the level of detail that users may read into them consciously or otherwise and subsequently draw conclusions that may or may not be accurate.

Implications for Digitally-Mediated Empathy

Our finding that the addition of EDA information resulted in significantly decreased EA performance ran counter to the expectation that FoV provided the least information, and would therefore perform worst in the rating task (see H1). On the contrary, EDA seemed to distract participants even when many reported ignoring it, and some described a mechanism by which the EDA signal prompted them to second-guess their ratings. Even with the presence of a first-person point of view as context, EDA was interpreted very differently among observers and thus is likely not a reliable cue if designing for observer accuracy, at least in the form explored here. This echoes qualitative findings from the related social biosensing research previously discussed, that EDA is often perceived as an unfamiliar and ambiguous signal for non-experts. One way to improve this may be to include a form of observer training in which they can see how their own EDA is affected by their emotions and thus establish improved intersubjectivity regarding the EDA information they receive about a target.

Though an increase in EA was expected and found for the high-fidelity, interpretable, and familiar information type of narrative text, it is unknown how dependent how reliant this type of cue may be on a target’s verbal expressivity. Less verbally expressive targets may find that biosensory information or other cues are comparable or advantageous compared with describing their emotions; the effectiveness of different informational cues for targets with a range of verbal expressivity is worth future exploration. In addressing H2, no significant effect of information type on rating error was found, suggesting that additional narration information did not improve observers’ accuracy of the magnitude of the target’s ratings over FoV, and neither did additional EDA worsen this measure of performance. Contrary to H3, information type had no significant effect on participants' state empathy. This may be because FoV alone already affords a high degree of feeling "in someone's shoes" leaving little room for improvement; participants' mean state empathy ratings were above 5 on a 7-point scale in all conditions.

We can compare these results to a similar study which assessed audiovisual, visual-only, and audio-only informational channels of targets describing emotional experiences. There, audiovisual information performed best at r = 0.47, followed by audio-only with r = 0.31, and lastly visual only with r = 0.21 [33]. In comparison, all of our EA values are relatively high, likely in no small part due to the content being available to observers and this content being a short film with an intended interpretation. The significant increase in EA in Narration and decrease for EDA however indicate that the content alone was not solely determinant of EA, and could be increased or decreased with the addition of supplementary information.

Two very different types of supplementary information's shared function as a reminder to consider the feelings of the target as separate from one's own was a particularly interesting finding. We saw this qualitatively in the interviews with observers and in the mean observer ratings in Figure 4 which centers around neutral more than in other conditions, an indication perhaps that an effect of the EDA information was that observers second-guessed their ratings and avoided the extremes of the scale in their predictions. Accuracy is neither wholly representative of real-life empathic processes, nor is it the sole important measure to track for improved mediated social interaction and understanding. The simple reminder to consider that a someone may be different from oneself with different reactions, feelings, and interpretations of a piece of content, a form of the aforementioned mind perception component of empathy, could be extremely valuable. A biosignal like EDA may allow this to be achieved anonymously removing the possibility to reinforce stereotypes or other prejudices, and can be produced automatically, an improvement over a cue like narration. Sharing experiences in this way could diversify and broaden observers' world views while preventing assumptions based on a target's identity.

5 LIMITATIONS

Scope Limitations

This study was conducted in a controlled laboratory setting, with a population of young adult US college students. Future work must consider cultural differences with more diverse populations and settings outside a research lab. We chose to have the target be unknown to observers, but in reality many mediated interactions occur between people who know information about one another to different degrees and at various levels of closeness. A public post or comment on Facebook for example typically includes a name and photo from which characteristics like race, age, sex, and nationality may be inferred (correctly or not). Further research should explore how this affects observers' motivation and ability to empathize.

Another limitation to this work is our use of a single target. While the purpose of this study was to test the effect of information channels on observer interpretations, different targets will have different reactions to the same content and it is important for future research to include a larger set of targets and with a range of narrative expressivity and EDA reactivity.

Operationalization of Empathy

The affective dial rating task method we used in this study is an artifact of the experimental design enabling quantitative analysis and testing hypothesis testing. In psychophysiology research EDA has been shown to be more related to the arousal dimension of affect than valence [27], and thus if the EA task was instead to rate the arousal of the results may be quite different which is worth future exploration. This and other work employing EA as a measure does so via valence rating accuracy because it is typically more salient than arousal in everyday interaction and interpretation of arousal may further complicates the already complex EA task. Still, the task is only a proxy for real-world empathy in which there are many varied informational cues available, and far more complexity than rating how positive or negative someone is feeling along a single dimension. Furthermore, the measures of empathic accuracy and rating error were quantifiable and thus enabled comparisons by condition, but are likely not the sole desired outcomes when considering someone else's point of view in an everyday scenario. We gathered qualitative feedback from our participants in semi-structured interviews because there is much to be learned even if the empathic accuracy results are ignored altogether; in a certain light the task itself can be seen as an interactive probe in which participants had some reason to focus their attention on understanding the original viewer's point of view, providing an experience to reflect on immediately after in the interview.

Future Work

The study of mediated empathy, especially the association with sharing biosensory information, is a young field and there remain many directions left to pioneer. This study in particular is suggestive of a few key pathways to pursue next.

Many participants reported it was easy to interpret the short film content we used in this study. To examine situations when empathizing may be more difficult (but likely more important), future work should test with more controversial and/or personally significant content for a target, when they may respond very differently than an "average viewer" or in a non straightforward way. Understanding differences between one another is as important for empathic processes as learning about similarities. Alternative content may not be entertainment-focused at all, and instead could be generated from participants' everyday experiences and thus be a higher fidelity representation of the type of content shared in mediated environments like social media platforms.

Another enlightening manipulation would be to vary the authority or certainty with which instructions regarding a signal like EDA are presented. This study used an ambiguous description, but it's important to ascertain the degree to which participants weigh unknown data types when empathizing if they are given an impression that it is a strong, accurate signal for defined emotional responses.

As described in the introduction, the analytical nature of the visual graphical display format of the target's EDA was expected to primarily affect cognitive empathy which this study focused on. Affective empathy, the direct feeling of another's emotions, and the effectiveness of different informational cues to produce it is worth future exploration. It is likely that in order for biosensory information to induce affective empathy, its representation should be more akin to the source of the signal as a physical manifestation rather than a visual number, icon, or graph.

6 CONCLUSION

Investigating the effects of different cues on the cognitive component of mediated empathy, we found that while narrative information resulted in higher accuracy inferring the valence of an anonymous other's experience, the addition of biosensory information significantly lowered accuracy echoing findings in related work that this information type is ambiguous and difficult to interpret precisely. Qualitatively we found that both types of information served as reminders of another's presence and that their experience may be different from one's own, suggesting that biosensory information, which can be collected and embedded automatically, can aid in the mind perception precursor to empathy. Designers of mediated communications systems should carefully consider these types of informational cues and their effects on users' ability to empathize with each other, especially given the ever-increasing prevalence and varied usage of mediated communications. Future work should continue to explore this and other empathy-related phenomena with larger and more diverse populations, in more diverse settings, and other creative designs implementing biosensory cues.

A SEMI-STRUCTURED INTERVIEW QUESTIONS

- In general, what did you think of what you did today?
- How did you split your attention between the video and [sensor/narrative] information?
- What did positive and negative ends of the scale represent for you?
- How do you think the sensor information matched what the viewer was feeling?
- What strategies or information did you use when predicting the viewer's feelings?
- How would you describe the viewer you rated?
- If you had to guess, what kind of person do you think the viewer is?
- How did you feel about the [text descriptions/sensor data] from the viewer?

ACKNOWLEDGMENTS

The authors thank Coye Cheshire and Noura Howell for their feedback on this manuscript, the staff at the UC Berkeley Xlab for their help recruiting and scheduling study participants, and to the participants themselves. The authors would also like to thank the CHI reviewers for their valuable comments and suggestions. This work was supported by a research gift from Facebook.

REFERENCES

[1] Sun Joo Ahn, Amanda Minh Tran Le, and Jeremy Bailenson. 2013. The effect of embodied experiences on self-other merging, attitude, and helping behavior. *Media Psychology* 16, 1 (2013), 7–38.
[2] Alessio Avenanti, Ilaria Minio Paluello, Ilaria Bufalari, and Salvatore M Aglioti. 2006. Stimulus-driven modulation of motor-evoked potentials during observation of others' pain. *Neuroimage* 32, 1 (2006), 316–324.
[3] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.
[4] Kirsten Boehner, Rogério DePaula, Paul Dourish, and Phoebe Sengers. 2007. How emotion is made and measured. *International Journal of Human-Computer Studies* 65, 4 (2007), 275–291.
[5] Nathan Bos, Judy Olson, Darren Gergle, Gary Olson, and Zach Wright. 2002. Effects of four computer-mediated communications channels on trust development. In *Proceedings of the SIGCHI conference on human factors in computing systems.* ACM, 135–140.
[6] Guillaume Chanel and Christian Mühl. 2015. Connecting brains and bodies: applying physiological computing to support social interaction. *Interacting with Computers* 27, 5 (2015), 534–550.
[7] Yawei Cheng, Ching-Po Lin, Chia-Yen Yang, Chao-Chih Wang, Daisy Hung, OJL Tzeng, Jen-Chuen Hsieh, Jean Decety, et al. 2007. The perception of pain in others modulates somatosensory oscillations. In *The Organization for Human Brain Mapping (OHBM) 2007 annual meeting (Chicago, USA: The Organization for Human Brain Mapping (OHBM)).*
[8] Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson. 2014. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication* 64, 2 (2014), 317–332.

[9] C Randall Colvin, Dawne Vogt, and William Ickes. 1997. Why do friends understand each other better than strangers do? (1997).

[10] Mark H Davis, Laura Conklin, Amy Smith, and Carol Luce. 1996. Effect of perspective taking on the cognitive representation of persons: a merging of self and other. *Journal of personality and social psychology* 70, 4 (1996), 713.

[11] Hillary C Devlin, Jamil Zaki, Desmond C Ong, and June Gruber. 2014. Not as good as you think? Trait positive emotion is associated with increased self-reported empathy but decreased empathic performance. *PloS one* 9, 10 (2014), e110470.

[12] William W Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 233–240.

[13] Hunter Gehlbach, Geoff Marietta, Aaron M King, Cody Karutz, Jeremy N Bailenson, and Chris Dede. 2015. Many ways to walk a mile in another's moccasins: Type of social perspective taking and its effect on negotiation outcomes. *Computers in Human Behavior* 52 (2015), 523–532.

[14] James J Gross and Oliver P John. 1995. Facets of emotional expressivity: Three self-report factors and their correlates. *Personality and individual differences* 19, 4 (1995), 555–568.

[15] Mariam Hassib, Daniel Buschek, Paweł W Wozniak, and Florian Alt. 2017. HeartChat: Heart Rate Augmented Mobile Chat to Support Empathy and Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2239–2251.

[16] Noura Howell, Laura Devendorf, Rundong Kevin Tian, Tomás Vega Galvez, Nan-Wei Gong, Ivan Poupyrev, Eric Paulos, and Kimiko Ryokai. 2016. Biosignals as Social Cues: Ambiguity and Emotional Interpretation in Social Displays of Skin Conductance. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. ACM, 865–870.

[17] William Ickes, Paul R Gesn, and Tiffany Graham. 2000. Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships* 7, 1 (2000), 95–109.

[18] William John Ickes. 1997. *Empathic accuracy*. Guilford Press.

[19] Joris H Janssen, Jeremy N Bailenson, Wijnand A IJsselsteijn, and Joyce HDM Westerink. 2010. Intimate heartbeats: Opportunities for affective communication technology. *IEEE Trans. Affective Computing* 1, 2 (2010), 72–80.

[20] Sara H Konrath, Edward H O'Brien, and Courtney Hsing. 2011. Changes in dispositional empathy in American college students over time: A meta-analysis. *Personality and Social Psychology Review* 15, 2 (2011), 180–198.

[21] Michael W Kraus, Stéphane Côté, and Dacher Keltner. 2010. Social class, contextualism, and empathic accuracy. *Psychological science* 21, 11 (2010), 1716–1723.

[22] Fannie Liu, Laura Dabbish, and Geoff Kaufman. 2017. Can Biosignals be Expressive? How Visualizations Affect Impression Formation from Shared Brain Activity. (2017).

[23] Fannie Liu, Laura Dabbish, and Geoff Kaufman. 2017. Supporting Social Interactions with an Expressive Heart Rate Sharing Application. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 77.

[24] Iris B Mauss, Robert W Levenson, Loren McCarter, Frank H Wilhelm, and James J Gross. 2005. The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion* 5, 2 (2005), 175.

[25] Hyeryung Christine Min and Tek-Jin Nam. 2014. Biosignal sharing for affective connectedness. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2191–2196.

[26] Jonathan W Peirce. 2007. PsychoPy - psychophysics software in Python. *Journal of neuroscience methods* 162, 1-2 (2007), 8–13.

[27] Jonathan Posner, James A Russell, and Bradley S Peterson. 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology* 17, 3 (2005), 715–734.

[28] Tania Singer, Ben Seymour, John O'doherty, Holger Kaube, Raymond J Dolan, and Chris D Frith. 2004. Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 5661 (2004), 1157–1162.

[29] Petr Slovák, Joris Janssen, and Geraldine Fitzpatrick. 2012. Understanding heart rate sharing: towards unpacking physiosocial space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 859–868.

[30] Petr Slovák, Paul Tennent, Stuart Reeves, and Geraldine Fitzpatrick. 2014. Exploring skin conductance synchronisation in everyday interactions. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. ACM, 511–520.

[31] Sophie Trawalter, Kelly M Hoffman, and Adam Waytz. 2012. Racial bias in perceptions of others' pain. *PloS one* 7, 11 (2012), e48546.

[32] Jamil Zaki. 2014. Empathy: a motivated account. *Psychological bulletin* 140, 6 (2014), 1608.

[33] Jamil Zaki, Niall Bolger, and Kevin Ochsner. 2009. Unpacking the informational bases of empathic accuracy. *Emotion* 9, 4 (2009), 478.