Essays on Mind-Reading Machines

Nick Merrill

November 8, 2017

1 Abstract

2 Optional preliminary pages

2.1 Dedication page

2.2 TODO Table of contents

2.3 List of figures, list of tables, list of symbols

2.4 TODO Preface or introduction

Machines will someday read the mind. I know. But they will. "We" will find a way to make them, and "we" will use any tools at our disposal to do so.

The primary tool we will use is our own interpretation, and theories of the mind, projecting it onto a physical reality that is less definite than "we" would sometimes like to imagine.

Who is this "we"? This dissertation looks primarily (though not exclusively) at software professionals in the San Francisco Bay Area.

etc. . .

Nick Merrill Berkeley, California

2.5 TODO Acknowledgements

3 How to sense a mind

This chapter outlines how theories of a mind that is embodied (beyond the brain), distributed (beyond the individual), and extended (to the built environment) make the mind amenable to sensing by emerging classes of wearable and environmental sensors.

3.1 OUTLINED Summary

- introduce mind reading & telepathy as terms for describing this goal
  - *mind*-computer interface (v. brain-computer interface)
- understand how tehoreis of mind support these terms
- recast past work under the guise of these terms
- surface the primacy of human interpretation, focus as a point of future (necessary) research

3.2 Mind reading & Telepathy

3.2.1 OUTLINED Summary

- introduce mind reading & telepathy as terms for describing this goal
  - *mind*-computer interface (v. brain-computer interface)

3.2.2 OUTLINED Telepathy

Telepathy
tele + pathos - Feeling / condition / state at distance.
Compare to

- *tele + vis* - Vision at a distance
- *tele + phon* - Hearing at a distance

Both would have been "magic" at one point!
Perhaps some technology will come along that also turns *telepathy* into a technical term.

3.2.3 OUTLINED Mind-reading

For our purposes,

- Telepathy: The **technical** aspects of encoding, transmission of mind
- Mind-reading: The **social** aspects, interpretation of encoded/transmitted mind

Some crap about how latter involves interpretation

However, we will see later in this chapter how unstable this "boundary" really is! Some crap about how theres no such thing as raw data etc

Some crap about how everything boils down to theory

3.2.4 OUTLINED Mind-computer interface

Mind-computer interface
Compare to BCI, why that term's not sufficient
Compare to affective computing, why that term's not sufficinet
Now, how to perofrm telepathy, mind-reading - in other words, how to produce an MCI? clearly we will need some theory to back us up.

3.3 Theories make the mind ammenable to sensing

3.3.1 OUTLINED Summary

this section, look at how existing theories help us in our quest to build an MCI show that this possibility is already very much here with current technology, doesn't require anything fancy and doesn't necessarily require improvements to brainscanning infra

3.3.2 **OUTLINED** The physical mind

What is it? What does it do? What is its relationship to the body? To the physical world? Two basic beliefs:

- **Dualism** - The mind has non-physical components
- **Physicalism** - The mind has only physical components

If we want to make mind amenable to sensing, we have only one choice! [1]

If we are **physicalists**, we believe some physical theory will eventaully explain the mind. Where do we look for such theories?

3.3.3 **OUTLINED** Cognitive science's physical mind

Cognitive science has historically been an influential source of such theories. Cognitive science's goal was and is a computational **model** of cognition. Much like physics can inform structural engineering, can these models inform neurosciecne [gallant], psychology [griffiths], AI [hawkins, minsky, bengio/ut guy], design [norman]. Compelling evidence for computational complexity explaining certain time-related tasks in psychological stimuli [e.g. Shepherd & Metzler, 1971] have inspired countless computational models, often in the form of flow diagrams that neatly compartmentalize cognitive "processes" such as long-term or short-term memory, attention, etc.

3.3.4 **TODO** Challenges to cognitive science

Cognitive science has received considerable criticism over its "isolationist assumtions." These critiques focus on two challenges, primarily.

- **Challenge 1**: Focus on the brain
- **Challenge 2**: Focus on the individual

1. **OUTLINED** Embodied cognition Challenge 1: Focus on the brain. Does brain == mind?

   (a) **Unstable brain/body dichotomy**. For example, neurons occur body-wide.

   (b) **Undervalues role of embodiement**.

   Response: Embodied cognition

   > "The exact way organisms are embodied simultaneously constrains and prescribes certain interactions within the environment."

   Resulting ultimately in the "Embodiement Thesis," that "**The agent's beyond-the-brain body** plays a significant causal role in that agent's cognitive processing." For example, offloading cognitive work onto the environment.

2. **OUTLINED** Distributed cognition Challenge 2: Focus on the invidiaul. As Clark et al propose, "...technological resources such as pens, paper, and personal computers are now so deeply integrated into our everyday lives that we couldn't accomplish many of our cognitive goals and purposes without them." Hutchins revolutionary analysis of pilots on a naval submarine brought this philosophical thesis into the empirical realm as "distributed conition."

3. **TODO** Activity theory

[1] Not necessarily the *only* choice! ...but the choice we will work with here. see Chalmers, D. (1996): The Conscious Mind. and Dennett, Daniel C. (1991). Consciousness Explained.

3.3.5 **OUTLINED** Making the mind ammenable to sensing

What good is all this theory? These three theories provide a picture of the mind that extends beyond the confines of the brain. As such, *these theories make the mind amenable to sensing without sensing the brain per se/make the mind amenable to sensing without sensing the brain per se.*

Now, if the mind is already *senseable*, has past work already sensed it?

3.4 Prior work on mind reading & telepathy

3.4.1 **OUTLINED** Summary

this section reviews past work, recasting it as being fundamentally *about* mind-computer interface. we make comparisons to the original author's aims, pointing out where and why our new lens on this work provides useful/fresh/new forms of analysis, and critique, as well as new directions for possible future work...

3.4.2 **OUTLINED** Affective computing

acknowledge that AC was onto the whole thing about emotions being sensible, being the domain of computers

concerned with mind-reading as *other people* do mind reading people who "mind reading" as term of art in autism, what autistics lack

1. **OUTLINED** Social Signal Processing (SSP) turn to easy critique of its view of emotions as computational, rather than co-constructed interpretive also easy critique of emotions being within a single person highlight alternatives e.g. boehner, leahu, sengers

2. **OUTLINED** Ground truth point out easy critique that it never found the ground truth it was 'looking for'

3. **OUTLINED** Affecting affect point out harder truth that it never looked much at how its judgement might AFFECT the phenomena it purports to observe

3.4.3 **OUTLINED** Pentalnd & distributed self

got closer with idea of distributed models like earthquake, perhaps emotions also distributed [parkinson]

easy critique that looked from perspective of manager, sought to make workspaces more efficient etc

harder critique that the CONSEQUENCE of this myopic focus missed some of the real concerns around privacy, autonomy, security

harder critique that this world has to some extent come to pass as a result targeted advertisements, digital surveillance to see whos terrorist, now with VR even more intimate bci promises yet another intimacy, look at those side-channel attacks and so on, done with P300 and now there's a startup with a P300 api [], once the stuff of fiction []

3.4.4 **OUTLINED** Transition

so i went through nitpicking all these theories, but how to remedy? what can we add to conversation that will substantially help to address these critiques across much research?

3.5 Human beliefs at the center of all

3.5.1 **OUTLINED** Summary

this section reviews some work that is substantially untouched by past work: the fundamental role of human beliefs in our ability to do anything with MCIs (technology in general) human beliefs always at some level of the stack. we also outline why this point matters, before concluding with a discussion of what can be done about it

3.5.2 TODO No fixed meaning a priori

3.5.3 TODO Material phenomena we *assign* meaning

3.5.4 TODO Why this matters

3.6 TODO Conclusions

people will continue to build increasingly hi-res models of human bodies in space;, and human environments I argued that these models will in general will be informative wrt *the mind*, producing what i dub *mind-computer-interfaces* (MCIs).

what 2 do now?

- It is now time to see if engineers believe mind is *senseable*
  - See how our theory matches up (or doesn't) with their beliefs
- One good starting place is the brain
  - But other wearable senosrs can also work .. heart is a good one, lots of connotations there, and may be diff btwn cultures!

WHY SO IMPORTANT?

4 Reading mind from heartrate

4.1 TODO list

4.1.1 TODO Mechanical stuff

1. **DONE** Add copy for ACM group
2. **TODO** Add figures for ACM group
3. **TODO** Add citations for ACM group
4. **TODO** Add copy for CSCW
5. **TODO** Add figures for CSCW
6. **TODO** Add citations for CSCW

4.1.2 TODO Tie together into chapter

1. **TODO** Fresh outline
2. **TODO** Draft relevant connective copy
3. **TODO** Copy/paste other text (use all text!)

4.2 ACM Group

4.2.1 Abstract

We investigate interpretations of a biosignal (heartrate) in uncertain social interactions. We describe the quantitative and qualitative results of a randomized vignette experiment in which subjects were asked to make assessments about an acquaintance based on an imagined scenario that included shared heartrate information. We compare the results of this experiment in adversarial and non-adversarial contexts of interaction. We find that elevated heartrate transmits cues about mood in both contexts, but that these cues do not appear to impact assessments of trustworthiness, reliability and dependability. Counter to our expectations, we find that normal (rather than elevated) heartrate leads to negative trust-related assessments, but only in an

adversarial context. Our qualitative analysis points to the role of social expectations in shaping contextual interpretations of heartrate, and reveals individual differences in the way interpretations are constructed. We unpack some of the ways that social meanings can arise from biosensor data, and discuss considerations for those designing interactions with wearables.

4.2.2 Introduction

What could your heartrate possibly tell another person about you? Past work has cited intimacy as a grounds for heartrate sharing [1,16,21,26]. However, it is not clear that these applications capture the full expressive capacity of heartrate as a computermediated cue. Does sharing a heartrate always "enhance" intimacy and trust? What is the social interpretation of an elevated heartrate in different contexts of interaction?

As biosensors become smaller and cheaper, more consumer devices are presenting users and application designers with opportunities to share their physiological data with others [18,22,23]. In fact, designers are already beginning to play with the notion that social and contextual interpretation may color people's interpretations of biosignals. The advertising copy for Cardiogr.am, one smartwatch app, reads,

> Your heart beats 102,000 times per day, and it reacts to everything that happens in your life—what you're eating, how you exercise, a stressful moment, or a happy memory. What's your heart telling you?

This application, along with many others, capitalizes on the idea that people project contextual interpretations onto heartrates, making them potentially interesting to share socially. However, few empirical studies examine how context shapes the interpretation of another person's heartrate [25], or how these interpretations might affect attributions of mood or emotion, or fundamental social attitudes such as trust.

In this paper, we investigate how individuals make social interpretations about a rudimentary biosignal (heartrate) in conditions of uncertainty, focusing on dyadic interactions between acquaintances. Dyadic relations, which are present in all groups, function as a fundamental starting point for understanding interpersonal collaboration and group interactions [4]. We describe the quantitative and qualitative results of a randomized vignette experiment in which subjects make assessments about an acquaintance based on an imagined scenario that included shared heartrate information. We examine two contexts in this study: an uncertain, non-adversarial context and an uncertain, adversarial context. These two contexts, differing only by a few words, ask participants to imagine they are meeting someone "for a movie" (non-adversarial) or "to discuss a legal dispute" (adversarial).

We find that a high heartrate transmits negative cues about mood in both contexts of interaction, but that these cues do not appear to impact assessments of trustworthiness, reliability or dependability. Counter to our initial predictions, we find that normal (rather than elevated) heartrate leads to negative trust-related assessments, but only in the adversarial context. In qualitative assessments of subjects' attitudes and beliefs, we find that normal heartrate in the adversarial condition conflicts with expectations about how the participant believes the acquaintance should feel, signaling a lack of concern or seriousness, which appears to lead individuals to view the acquaintance as less trustworthy. In contrast, subjects in the non-adversarial context relate elevated heartrate to empathy and identification rather than trustworthiness. We also find a small number of subjects read different social interpretations onto the heartrate signal, including a very small minority who did not infer any relationship between the heartrate and the social situation.

Where past work establishes that heartrate has contextual meanings, this study aims to establish the relative effect of the elevated (versus normal) signal on social attitudes regarding trust, reliability and dependability. Understanding these effects could shed light on why people agree or refuse to disclose biosignals in interpersonal relationships and groups, and what individuals risk by making such disclosures. As we better understand the social interpretations of biosignals, we can inform the design of technologies that facilitate biosignal sharing in groups.

4.2.3 Social interpretations from physiological signals

To date, most work on the contextual interpretation of sensor data has focused on individual interpretation of individual data (c.f. quantified self). In contrast, our work attempts to move toward an understanding of how biosignals are interpreted in interpersonal interactions – the quantified social self. This shift is motivated, in part, by an increasing number of consumer applications that support sharing biosignals such as heartrate. Especially pertinent to our study, it is not well understood what heartrate actually signals to another person in a social interaction. How might the contextual, social interpretation of another person's biosignals affect social interpretations of mood (e.g., anxiety, calmness), or attitudes about trustworthiness and dependability?

Goffman [12:56] makes an important distinction between the cues that we intend to give to others, and those that are "given off" unintentionally through our numerous non-verbal actions and behaviors. We view physiological signals such as heartrate as a form of non-verbal signaling that can "give off" more information to others than the sender may desire [15]. This type of personal data revealed through discreet sensors paired with mobile communication technologies has, until recently, been unavailable in most forms of social interaction.

As Donath (2007) argues, when we use communication technologies for interpersonal and group interactions, we do so as a way to keep track of our evolving relationships with other people and, critically, to know whom to trust [10]. Crucially, we rely on signals such as verbal communication, facial expressions, gestures, and other patterns of activity to help us fill the gap between what someone is doing and why they are doing it [11]. So, when new forms of biosensory information, like heartrate, are suddenly available in social interaction, significant questions arise surrounding the inferences that we might draw from these signals, how the context in which they are presented may shape our interpretations, and how this information may affect our subsequent beliefs and behaviors towards others.

1. Contextual interpretations of sensor data

   Prior work interrogates the contextual interpretation of personal data from certain kinds of sensors [6,7], but physiological data has received less attention, despite two crucial differences from sensors that capture information such as location (e.g., GPS). First, biosensor data are intrinsically ambiguous: whereas a GPS coordinate refers to one specific place, heartrates do not have oneto-one mappings to physical activities or emotions. Second, physiological phenomena vary from person to person; 60bpm could be high or low depending on whose heartrate it is. A relatively large body of work has looked at how the transmission of physiological data might play a role in computermediated communication. One class of application has attempted to explicitly encourage or discourage certain behavioral outcomes, making some biosignals apparent such that the transmission of the data acts as a social cue [2,19]. Another class of prototypes explores how signals might affect feelings of intimacy, particularly between romantic partners [1], and several applications focus on the transmission of heartrate as a means to achieve this effect [16,25,26].

   Despite a number of applications that transmit biosignals socially, there remains little work on how people interpret the biosignals they see from others. Past studies on heartrate sharing indicate that people do read social emotional cues in the heartrates of other people [20,25]. However, it is still not well-understood what social meanings these signals take on in different social contexts and, in contrast to the findings of previous studies, it is not clear that the social consequences of transmitting physiological data will always be positive (e.g., increased intimacy).

   In this study, we are particularly interested in trust, due its centrality to the study of group behavior, and to the claims of prior research on heartrate sharing [16]. We apply the encapsulated interest view of trust [14], in which one person's interest is intrinsically connected with the interests of another person [5]. Since trust between individuals is only salient in the presence of risk and uncertainty [8], we test our hypotheses about the effects of heartrate information on evaluations of trust in social contexts with different degrees of risk and uncertainty.

2. The social meaning of a heartrate

   Compared to social interpretations of physiological signals, interpretations of one's own signals are slightly better-understood from empirical research. Individuals' interpretations of their own heartrate have received particular attention (see [24] for a review). Studies have generally revealed that, when individuals believe that their heartrate is elevated, they sometimes believe their mood and emotions to be more negative [27].

If lay interpretations of one's own heartrate can yield negative self-interpretations, sharing heartrate information could also yield negative effects on mood and trustworthiness, particularly during uncertain interactions where something is at stake (such as time, money, or other valued resources). To investigate, we use a mixed-methods approach combining quantitative and qualitative analyses of a survey-based vignette experiment.

4.2.4 Interpreting biosignals in uncertain social interactions

Based on aforementioned studies of individual's negative emotional interpretation of their own heartrate, we believe that this negative valence will be mirrored in people's interpretations of the heartrates of others in uncertain situations. Our investigation begins with two key predictions about negative assessments of one's partner in an uncertain social situation. We test both hypotheses in two different contexts of interaction (adversarial and non-adversarial) to understand how the context of risk and uncertainty affects social interpretations of heartrate.

1. Heartrate and Mood Past work indicates that people tend to make negative inferences about mood and emotion from elevated heartrates [9,13,27]. As such, our first hypothesis predicts that participants will adjust their attitudes about the mood of their partner when their partner's heartrate is elevated, as opposed to normal: Hypothesis 1: When individuals believe that their partner has an elevated heartrate in an uncertain social interaction, they will report their partner as being (1a), less calm (1b), more emotional (1c), and more easily upset (1d), compared to those who believe that their partner has a normal heartrate.

2. Heartrate and Trustworthiness Where Hypothesis 1 predicts that individuals will make negative assessments about an acquaintance's mood based on elevated heartrate, our second hypothesis predicts that individuals will make negative assessments about dispositions to behave in a reliable, dependable and trustworthy manner. Thus, both hypotheses stem from the same base assumption that, all things being equal, elevated heartrate has a primarily negative connotation with attitudes and behaviors of another person. Hypothesis 2: When individuals believe that their partner has an elevated heartrate in an uncertain social interaction, they will make negative assessments about the partner's trustworthiness (2a), reliability (2b), and dependability (2c), compared to those who believe that their partner has a normal heartrate.

4.2.5 A survey-based experiment

To test our hypotheses, we conducted a survey-based vignette experiment. Vignette studies involve short descriptions of a scenario, designed to elucidate opinions, attitudes, and beliefs about that particular situation [17].

In this vignette study, we compare two different contexts of interaction. We do not create separate hypotheses for the two different contexts; rather, we are interested in comparing and contrasting the two different contexts to see how they might interact with social interpretations of heartrate. We provide our participants with either an adversarial or a non-adversarial social context. In the adversarial scenario, the participant is waiting to meet an acquaintance about a legal dispute. In the non-adversarial scenario, the participant is waiting at a movie theater for an acquaintance so that they can see a film together.

In all scenarios, the acquaintance sends a message via smartphone indicating that he or she is running late due to slow traffic. The person who is waiting does not know if the acquaintance will make it on time or not, or whether the acquaintance is being honest about their tardiness. Within each context, we manipulate a small piece of information about the heartrate of the acquaintance: We tell the participant that the heartrate of the acquaintance has been shared by the acquaintances' smartphone and it is either elevated or normal.

1. Sample

   Our sample was undergraduate students recruited from the population of a large, public university on the West Coast of the U.S. Potential participants were asked to participate in a short online survey, and they did not know the nature of the questions or the topic of the study in advance. All participants were paid a $5 Amazon gift card. One hundred and three participants (103) completed the experiment survey instrument. The pool was weighted toward women; in our sample, 65% were women and 34% are male, and 2% (2 subjects) did not identify with either gender. With random assignment, the same overall gender split was maintained across conditions. The mean age of participants was 23.

2. Vignettes Each participant in the study saw only one of the four possible vignettes. After the vignette, the survey included free response questions about subjects' reactions to and interpretations of the situation described in the vignette, as well as 7-point Likert scale questions (Strongly Agree to Strongly Disagree) in which subjects evaluated the other person's disposition ("This person is emotional", "This person is anxious", "This person is easily upset", and "This person is calm"). In addition, we asked participants to indicate whether the other person was "trustworthy," "reliable," and "dependable" using the same 7- point agreement scale.

   There are two contexts of interaction (adversarial and nonadversarial) and two heartrate conditions (normal and elevated), creating four distinct vignettes based on social context and heartrate (HR): adversarial elevated HR, adversarial normal HR, non-adversarial elevated HR, and non-adversarial normal HR. Participants were randomly assigned into one of the four conditions. We manipulated these heartrate conditions by making a key wording change as indicated in the two context vignettes below.

   (a) Non-Adversarial

   > You planned to meet your acquaintance for a movie at seven. It's 7:15, and you're standing alone in front of the theater. Your phone buzzes, and you see a message from this person that says, "I'm running late, traffic was really slow." Through your smartphone, you are able to see this person's heartrate, which the app designates as [normal / elevated]. It is currently 75 degrees and sunny. Your movie starts at 7:20.

   (b) Adversarial:

   > You planned to meet your acquaintance at seven to discuss a difficult legal dispute between the two of you. It's 7:15, and you're standing alone in front of the meeting spot. Your phone buzzes, and you see a message from this person that says, "I'm running late, traffic was really slow." Through your smartphone, you are able to see this person's heartrate, which the app designates as [normal / elevated]. It is currently 75 degrees and sunny.

3. Quantitative results

   We apply both quantitative and qualitative analyses to investigate our research questions and hypotheses. The study is based around an experimental design, but we also place significant emphasis on open-ended responses to better understand participants' thought processes, beliefs, and rationale for their choices in the vignettes. Our first hypothesis predicts that individuals will make negative attributions about the mood of the acquaintance in this uncertain situation when they believe that the acquaintance has an elevated heartrate (compared to normal heartrate). Given our four separate measures of mood, we conducted a multivariate analysis of variance (MANOVA) to test the hypothesis that there are one or more mean differences between the normal/elevated heartrate conditions, and/or between the two contexts of interaction (nonadversarial and adversarial).

   We found a strong, statistically significant effect and a medium practical association between emotional attributions and heartrate condition, $F\ (4, 96) = 32.89$, $p < .001$; partial eta squared = .58. Turning to the individual outcomes, we find that subjects' perceptions of the acquaintance in the vignette's anxiety, his/her tendency to be easily upset, his/her tendency to be emotional, and his/her lack of calmness were all significantly higher in the elevated heartrate conditions when compared to the normal heartrate conditions (see Figure 1). We found no significant effect for the two contexts of interaction, $F\ (4, 96) = 1.072$, $p = .38$, and no significant effect for the context x heartrate condition interaction, $F\ (4, 96) = 1.65$, $p = .17$. In sum, individuals significantly rate acquaintances with elevated heartrate as more anxious, easily upset, and less calm than those with normal heartrates. In the non-adversarial context, individuals did not rate the acquaintances as significantly more emotional in the elevated condition compared to normal, but this difference was statistically significant in the adversarial context.

The context of interaction (non-adversarial, adversarial) does not have any effect on mood ratings. With clear statistical and practical significance for the overall effect of mood attributions by heartrate condition in both contexts of interaction, Hypothesis 1 is supported.

Our second hypothesis predicts that individuals will make negative assessments about how certain they are regarding the acquaintances' trustworthiness characteristics when the individual has an elevated versus a normal heartrate. We find a statistically and practically significant effect for the heartrate conditions, F (3,

(a) $= 4.19$, $p < .01$; partial eta squared $= .12$. However, we also

find statistically significant effects for both the context of interaction, F (3, 97) $= 2.82$, $p < .05$, and the context x heartrate condition interaction, F (3, 97) $= 2.75$, $p < .05$.

A closer inspection of the individual mean differences reveals that the means for all three outcomes (reliability, dependability and trustworthiness) are all lower in the normal condition compared to the elevated condition in the adversarial context (see Figure 2). This result is the opposite of what Hypothesis 2 predicts. In the non-adversarial context, we find no statistically significant differences in trust-related evaluations between heartrate conditions. Thus, it is the interaction between the context and the heartrate condition that explains the results: individuals rate acquaintances with normal heartrates significantly lower in terms of trustworthiness, dependability and reliability than those with higher heartrates—but only in the adversarial condition.

Individuals do not rate acquaintances any differently on these three outcomes between the heartrate conditions within the nonadversarial context. In fact, the means for these outcomes are very similar across all conditions and contexts, with the sole exception of the adversarial, normal condition. The mean differences for the trust-related outcomes between the normal and the elevated conditions within the adversarial context are all highly statistically significant ($p < .01$) and highly practically significant: Cohen's d $= 1.1$ (trustworthiness); 1.07 (dependability); 0.68 (reliability). Hypothesis 2 is therefore not supported. However, the strong findings (statistically and practically significant) in the opposite direction from our prediction warrant further exploration in the qualitative results and discussion below.

4.2.6 Qualitative results

Directly after the vignette, participants were asked four freeresponse questions about their reactions to the situation described in the vignette: 1) How do you react to this message, 2) What makes you react this way, 3) What is the ideal outcome of this situation, and 4) What is the worst possible outcome of this situation? The open-field responses were coded into two broad, non-overlapping categories: those that mentioned a negative emotional reaction to the scenario, and those that included a mention of what the other person in the situation might be thinking or feeling. Responses in the latter category were further sub-divided by experimental condition for analysis.

1. Adversarial Context

   This section reports on the qualitative analysis of free responses given by subjects in the adversarial (legal dispute) context.

   (a) **TODO** Normal heartrate

   <u>Figure 1. Mood-related evaluation means by condition (bars represent standard deviation).</u>

   <u>Figure 2. Trust-related evaluation means by condition (bars represent standard deviation).</u>

   In the adversarial (legal dispute) context, many subjects who saw a normal heartrate directly indicated that they were negatively adjusting their appraisal of the other person, either in their sympathy toward the other person, or in their judgment of that person's trustworthiness. We find that normal heartrate in the adversarial condition appears to be in conflict with the subjects' expectations about how the acquaintance should feel.

> I will feel less sympathetic to this person because their heart rate doesn't show that they are stressed or upset.
>
> I feel annoyed because a higher heart rate would indicate that the person cares about the meeting
>
> The normal heartrate implies that my acquaintance isn't taking this meeting seriously. However, it is difficult to say that my acquaintance does not care or is lying. For example, I have no knowledge of the traffic to determine if my acquaintance is lying.

Here, participants read a lack of care or concern into the acquaintance's normal heartrate, but did not feel the biosignal provided definitive evidence as to whether or not the acquaintance was being truthful. For some participants, however, normal heartrate indicated deception:

> I would think this person is lying. If they were in a rush, their heartrate would be faster.
>
> I feel like he is lying and is taking his time. I say "hurry up please I can't wait any longer. You are lying to me" It makes me angry to see that his heartrate is normal through all of this. Mine is spiking out of control.

These responses could help to explain the surprising quantitative results of Hypothesis 2 in the adversarial context: the intersection of the adversarial context with normal heartrate led many participants to view the acquaintance as unsympathetic and, in some cases, disingenuous. As we see below, these negative reactions stand in stark contrast to the interpretations in the elevated heartrate condition.

(b) Elevated heartrate

In general, participants in the adversarial context viewed elevated heartrate as a signal that the acquaintance cared about being late.

> Since it shows that the person is trying their best to come, as shown by the elevated heartrate, I would still feel ok.
>
> I would believe my acquaintance. An elevated heartrate tells me she is probably rushing/hurrying over. I have data from the phone to validate what she is saying to a certain extent.

In these quotes, participants used the elevated heartrate to validate their acquaintance's claim, thus positively assessing their honesty. A few subjects spoke to the power of data in creating what appeared to be objective facts about the other person.

> I won't be angry because seeing this person's heart rate being elevated, it must mean they're in a hurry. Seeing metrics make it easier to believe someone.
>
> I feel like I'm in a position of power. With the capacity to check someone's heart rate, I can instantly tell how they are feeling. In a way, it is almost like a lie detector.

In both of these quotes, we see attitudes about the presumed authority or "neutrality" of data interacting with beliefs about the body (namely, the relationship between heartrate and emotion, or truthfulness), creating a context in which wearables data can be used to construct social judgments or assessments. How these assessments play out will vary in different social situations, with different sensors, and in different contexts of use. Such variations should be explored much more deeply in future work.

2. Non-Adversarial Context

This section reports on the qualitative analysis of subjects in the non-adversarial context (meeting for a movie),

(a) Normal heartrate

In the non-adversarial context, many participants reported that normal heartrate conveyed a lack of appropriate social concern:

> At first I believe that maybe my acquaintance is running late; however, when I discover that their heart rate is normal I wonder why it isn't higher...
>
> It seems like they are too nonchalant about it
>
> I feel frustrated because it seems like the person isn't concerned about making me wait.

In these cases, interpretations focused on what the other person was thinking or feeling. As we saw in the adversarial context, normal heartrate seems to be in conflict with expectations. However, unlike in the adversarial context, we did not find evidence that subjects were re-appraising their trust toward the other person. Interestingly, two participants read the normal heartrate positively, as a sign that the other person was telling the truth.

> If his heartrate is normal, then he is probably not lying. I would still be slightly annoyed at this.
> it's OK. her heartbeat was normal, so no lies

These subjects seemed to feel annoyed by the partner's normal heartrate. However, in contrast to the adversarial context, no subjects explicitly stated that the other person seemed less trustworthy, honest or reliable as a result.

(b) Elevated heartrate The majority of respondents in the non-adversarial indicated that the elevated heartrate was a token of the other person's regret for being late to the movie. Many participants in this condition indicated that they would have a more sympathetic reaction to the text message as a result of seeing an elevated heartrate.

> Elevated heart rate tells me that the acquaintance at least cares that he/she is late and there's no point in getting mad.
> I would text her back "No problem! I'll grab the tickets and will wait for you out front." It seems obvious she's in a hurry to get there, and is late because of traffic.
> I will feel apologetic because I can see that this person's heartrate is elevated and I do no want him/her to feel worried/ stressed about making a movie.
> I would feel anxiety about being late for the movie and pity because they seem anxious.
> I don't like being rushed and get anxious when I am rushed

In these responses, heartrate generally seemed to signal that the acquaintance was stressed. While stress is generally assumed to be negative, in this case it seems to engender identification and empathy with the acquaintance. This example gestures toward the highly contextual nature of heartrate's social meaning, and why more work should examine the consequences of these different interpretations.

3. Other interpretations of heartrate: Relevance, validity, creepiness

In addition to the major themes noted above, we also found a few other important interpretations. A small handful of participants (12 total) mentioned aspects other than the immediate social interaction in relation to the shared heartrate display. The points that surfaced surrounded concerns about privacy, doubts about the accuracy of the sensing device, and doubts about the relevance of heartrate to the particular context.

(a) Privacy and disclosure concerns

Only three subjects in the entire experiment pool (n=103) commented on the potential for invasiveness or over-disclosure in heartrate sharing.

> (non-adversarial + normal heartrate) "I feel like I'm violating my acquaintance's private information by knowing their heart beat."
> (adversarial + normal heartrate) "I do suspect the person is lying since his heart rate is normal. I think the extra info of the heart rate is the reason I have a neg. suggestion towards the person. I think the reported heart rate is a bad idea."

Given that heartrate sharing is not (yet) widely deployed in consumer devices, it is somewhat surprising that only a few subjects commented on privacy concerns. This could be partially explained by the fact that the scenario was imagined, rather that simulated, and because subjects might have anticipated our interest in their reactions to the interface.

(b) Validity of the device's data

Four subjects mentioned the possibility that the device, or the intuitive inferences drawn from it, may be inaccurate. (adversarial + elevated heartrate) Heart rate could be elevated for many reasons, and just like studies with lie detectors, it may possibly indicate lying, but also could indicate other things. It's just a number, not a definite answer of lying or not. And even then, you've got to forgive people.

> (adversarial + normal heartrate) "The normal heartrate implies that my acquaintance isn't taking this meeting seriously. However, it is difficult to say that my acquaintance does not care or is lying. For example, I have no knowledge of the traffic to determine if my acquaintance is lying. Additionally, my smartphone can be wrong; I don't know how accurate this technology is, especially since it is a very new piece of technology."

Our study did not reference any existing device, so it is possible that the fallibility of particular devices was not on subjects' minds. However, the trust that people place in sensing devices, and the presumed authority of their data, should be explored thoroughly in future work.

(c) Relevance of heartrate to the social situation

Only two subjects in the study who mentioned heartrate felt that the data was not necessarily related to the specific social situation described in the vignette:

> (non-adversarial / elevated heartrate) "My initial reaction would probably be to ask them if everything is okay. Their heart rate should probably not be elevated since they are only driving and weather conditions are not abnormal."
> (adversarial / normal heartrate) "There may be reasons why his/her heartrate is normal and why he/she may be late in the first place, so I'm not concerned about that."

Across all conditions, the fact that the vast majority of participants inferred a causal relationship between the heartrate information and the particular social situation highlights the relatively reliable effect of context in priming subjects to draw such inferences. Our results indicate that simply making the heartrate salient, in the absence of other cues, invites people to project a causal narrative on the mood, intentions, and behavior of others.

4.2.7 Discussion

We began this investigation by asking how individuals might interpret heartrate information in uncertain social interactions. Our hypotheses are both based on the simple rationalization that the kinds of negative attributions that people tend to make about their own heartrate will be echoed in their social interpretations of others' heartrates in uncertain contexts. We found, however, a much more complex story about the social interpretation of biosignals and the context of interaction.

Our first hypothesis predicts that an elevated heartrate will be negatively associated with assessments about mood and dispositions in uncertain social interactions, both adversarial and non-adversarial. We found strong support for this hypothesis in both contexts, across our outcome attributions, in line with prior works' findings regarding interpretation of one's own heartrate [24]. Our second hypothesis predicts that an elevated heartrate will lead to negative assessments about the partners' trustworthiness, dependability and reliability. As with our first hypothesis, we expected that pre-existing negative connotations with heartrate might translate into negative expectations of trustrelated behavior.

We rejected the second hypothesis in both contexts of interaction. In the non-adversarial context, we found no difference in assessments of trustworthiness, dependability or reliability in the elevated and normal heartrate conditions. Furthermore, we found that the average assessments on these three outcomes were nearly identical between the elevated condition in the adversarial context and the elevated and normal conditions in the non-adversarial context.

Most surprisingly, we find a decrease in trustworthiness, dependability, and reliability in the normal heartrate condition, but only in the adversarial context. As noted in the quantitative results, the differences between the elevated and normal conditions in the adversarial context were highly statistically significant: each of the trust-related measures saw an average decrease of one full point (on a 7-point scale) in the normal condition compared to the elevated condition.

To help explain these results, we turn to our qualitative analyses of the adversarial (legal dispute) context. Subjects in the adversarial context seemed to have expected their partner to have an elevated heartrate. When the partner had a normal heartrate, participants viewed it as evidence that s/he is not bothered enough, not taking the situation seriously, or perhaps even lying. Indeed, many participants explicitly stated in the open text responses that they trusted the partner less because his or her heartrate was normal.

Why do we not see the same effect in the non-adversarial context? Turning again to the qualitative data, we find that participants took elevated heartrate as a token of their acquaintances' genuine desire to arrive on time. It seems that elevated heartrate led many participants in the non-adversarial context to increase their empathy, identification, and understanding of the partners' situation. Thus, even though individuals in the non-adversarial condition associate elevated heartrate with anxiety, lack of calmness, and being easily upset, the negative emotional interpretations do not seem to translate to evaluations of one's trustworthiness, dependability or reliability.

Taken together, we see that heartrate does not inherently (or consistently) affect trust-related outcomes. Instead, social expectations shape interpretations of the heartrate biosignal to create highly contextual, socially-specific meanings. CMC researchers have long noted that, when cues are omitted from technology-mediated interaction, people tend to fill in the gaps [3,10]. However, individuals may interpret new types of interpersonal data in ways we do not yet understand. Our work provides some evidence that such interpretations might have real social consequences. The fact that heartrate alone can significantly alter one's perception of trustworthiness in an adversarial context is an important step towards the larger goal of unpacking social interpretations (and their effects) in technologysupported social interaction. (For one thing, the mostly positive social interpretations of heartrate observed in past work are likely highly dependent on the social context in which they were observed).

Finally, we note a diversity of opinions and interpretations within conditions. For example, a few subjects took normal heartrate as proof of honesty, the opposite view from the majority of subjects. A few subjects did not feel there was necessarily any relationship between heartrate and the social situation at hand. A small minority (three subjects) mentioned concerns around privacy or disclosure. The wide range of views, sometimes contradictory, highlights the complexity intrinsic to interfaces that collect and share biosignals, and warrants future studies into social and contextual interpretation of data from wearable devices.

4.2.8 Limitations

Our vignette experiment examined a single type of scenario in two different contexts, using text-based answers. We still have a limited picture of the range of theoretically important contexts in which individuals may observe and interpret biosignals about others, and a limited understanding of how the rich cues present in realistic interaction contexts might bare on our findings. Our study focused on a first-time interaction with an imagined heartrate sharing interface. We do not know how our findings would hold over time, and it is very likely that social meanings of any biosignal could become more consistent over time. The vignette scenario was contrived from believable, but currently non-existent smartphone technology. Either due to participants' suspension of their disbelief or due to their actual attitudes about the heartrate sharing, few participants raised questions regarding privacy implications of these scenarios.

Since the vignette study took place online, we could have missed the sorts of rich contextual cues that might be captured by live interviews or other in-person methods. Furthermore, the internet presents a wide array of distractions to survey-takers, and our survey was not able to detect the participants' attention on the task (e.g.., we could not detect whether the subject was switching between tabs in their web browser, or taking breaks during the survey), nor did we monitor how long subjects spent filling out the survey.

4.2.9 Future work

Our experiments provide evidence that interpretations of biosignals from sensors (such as wearables) can affect social attributions and behaviors towards others. Nevertheless, many questions persist. Future work should approach similar topics with controlled, lab-based experimental methods. Controlled, behavioral experiments could help us ask more specific questions about how elevated heartrate affects perceptions of risk in uncertain interactions, e.g. when money is at stake. Such studies could lead to a more robust understanding of how the transmission of biosignals might affect social behavior. Future work should also investigate the degree to which our findings are unique to heartrate. Other biosignals (e.g., galvanic skin response, electroencephalography or EEG), could offer different affordances for sense-making. It is unclear from our work how the social interpretation of the signals from these devices could affect social behaviors such as dyadic and group trust.

Future work should also investigate the mechanisms by which contextual interpretations of heartrate arise. Specifically, future work should study the effect of the signal's representation on the attributions made by participants. In this study, subjects saw a normative interpretation of heartrate ("elevated" versus "normal"); however, it is unclear how a more more or less ambiguous, or more or less numerical interface might have affected our results. Meanwhile, the meaning of a given biosignal is certainly mediated by a number of factors, including culture and lived experiences. Much work remains in determining how biosignals give rise to any interpretations, and to what extent these interpretations are influenced by both beliefs about a given signal and the way interfaces represent this signal.

Finally, future work should more closely examine the relationship between people's social interpretations of biosensor data, and their assessment of the risk associated with disclosing these signals. As biosensing devices become more pervasive, people will more often be faced with decisions to disclose these data to others, either socially, or for services (e.g., health care, insurance) [23]. Consumers' willingness to (or refusal) to disclose their signals will certainly stem from their intuitive interpretation of what these signals might say about them, at least in part. As such, processes of consent in disclosing wearables data present a clear priority for future work.

4.2.10 Conclusion

We find that sharing heartrate can negatively influence trusting attitudes. However, heartrate alone does not communicate trust. Instead, individual's social expectations interact with the heartrate data to produce context-specific meanings. Complicating matters further, our qualitative data reveal a diversity of interpretations regarding the relevance and meaning of a heartrate in context, and the privacy implications of biosensing technologies. Our findings advance and complicate our understanding of the role that biosignal sharing can play in social, computer-mediated contexts, and motivate more detailed study into the mechanisms by which social interpretations arise from basic physiological signals.

5 Shifting to the brain

6 Talking to engineers about brain-computer interface

7 Who are you really? Probing engineers on authentication and the ground truth of identity

8 Beliefs, practices

9 References

10 Bibliography

11 Appendices