Title

Mind-Reading and Telepathy for Beginners and Intermediates: What People Think Machines Can Know About the Mind, and Why Their Beliefs Matter

Permalink

Author

Merrill, Nick

Publication Date

2018

Peer reviewed|Thesis/dissertation

Mind-Reading and Telepathy for Beginners and Intermediates: What People Think Machines Can Know About the Mind, and Why Their Beliefs Matter

by

Nicholas Julius Merrill

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Information Management and Systems

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John Chuang, Chair
Associate Professor Coye Cheshire
Associate Professor Alva Noë

Summer 2018

Mind-Reading and Telepathy for Beginners and Intermediates: What People Think Machines Can Know About the Mind, and Why Their Beliefs Matter

Abstract

Mind-Reading and Telepathy for Beginners and Intermediates: What People Think Machines Can Know About the Mind, and Why Their Beliefs Matter

by

Nicholas Julius Merrill

Doctor of Philosophy in Information Management and Systems

University of California, Berkeley

Professor John Chuang, Chair

What can machines know about the mind, even theoretically? This dissertation examines what people (end-users and software engineers) *believe* the answer to this question might be, where these beliefs come from, and what effect they have on social behavior and technical practice. First, qualitative and quantitative data from controlled experiments show how basic biosignals, such as heartrate, meet with both social context and prior beliefs about the body to produce mind-related meanings, and affect social decision-making. Second, a working brain-computer interface probes the diverse beliefs that software engineers hold about the mind, and uncovers their shared belief that the mind can and will be read by machines. These cases trace an unstable boundary—one heavily mediated by human beliefs—between sensing bodies and sensing minds. I propose the porousness of this boundary as a site for studying the futures of computer-mediated communication, of security, privacy and surveillance, and of minds themselves.

To Mom

I've been shaving (mostly). Thank you for everything. I love you forever.

Contents

List of Figures

List of Tables

Acknowledgments

Thank you to John Chuang for being a mentor in every sense of the word, for accepting me into this ancient model of apprenticeship and showing me how the work is done. During my time as a PhD student, John assembled BioSENSE, a group that made my dissertation possible. A few BioSENSors in particular helped me immensely: Coye Cheshire (with whom I did much of the work in this dissertation), Richmond Wong, and Noura Howell. Outside of BioSENSE, Alva Noë raised the relationship between my work and related debates that were hiding in plain sight. Along the way, Paul Duguid taught me a great deal about what good scholarship looks like. I would like to thank all of my collaborators, inside and outside the ISchool, and to thank the ISchool itself for providing me the freedom and autonomy to pursue a risky topic.

Above all, I would like to thank Mom and Dad, who always believed in the importance of my work, even when they weren't quite sure what I was doing.

I would also like to thank Chihei Hatakeyama and Hakobune for the peaceful music.

Chapter 1

Introduction

What can machines know about the mind? This dissertation seeks to understand *people's beliefs* about this question: how these beliefs affect and arise from interactions with digital sensors, from prior beliefs about the mind and the body; and how these beliefs may shape the design of technical systems in the future.

The purpose of this dissertation is twofold. First, it surfaces that the boundary between sensing bodies and sensing minds is unstable, deeply entangled with social context and beliefs about the body and mind. Second, it proposes the porousness of this boundary as a site for studying the role that biosensing devices will play in near future. As biosensors creep into smart watches, bands, and ingestibles, they will build increasingly high resolution models of bodies in space. Their ability to divine not just what these bodies do, but what they think and feel, presents an under-explored avenue for understanding and imagining how these technologies will come to matter in the course of life.

Chapter 2 begins by introducing the notion that the mind is readable from consumer devices worn on the body and embedded in the environment. It reframes some past studies in computer science and adjacent fields as having already begun the work of theorizing and building computational *models of minds* (Section 2.2). It then motivates human beliefs as a starting point for discovering the relevance of the readable mind, both in how engineers will model it, and how end-users will encounter these models in life.

With focus fixed on human beliefs, Chapter 3 describes an empirical examination of how people conceive of the mind with respect to heartrate, a popular sensing modality in commercial devices. Through a vignette study, this chapter demonstrates that heartrate can take on various, sometimes contradictory meanings in different social contexts.

While this study establishes that people can build mind-related meanings around basic biosignals, it does not establish whether these beliefs can affect social behaviors, nor how specific our findings are to heartrate. In Chapter 4, we apply quantitative and qualitative analyses to an iterated prisoner's dilemma game, in which heartrate information ("elevated" or "normal") was shared between players. In a follow-up study, we replicate our initial study, but replace heartrate with an unfamiliar biosignal, "Skin Reflectivity Index (SRI)." We find that both heartrate and the unfamiliar biosignal are associated with negative mood attributions

when elevated, but we observe a decrease in cooperative behavior only with elevated heartrate. Our findings highlight the role beliefs about the body can play in shaping interpretations of a biosignal, while simultaneously suggesting that the social meaning of unfamiliar signals can be "trained" over repeated interactions.

The prior two chapters establish that the mind-related meanings of biosignals, familiar and unfamiliar, arise from both social context and prior beliefs about the body. But how do the basic biosignals we studied compare to the wide variety of sensing modalities emerging in consumer devices? Chapter 5 explores beliefs about a variety of biosensing devices, examining how people relate their data to qualities of mind. I report on the qualitative and quantitative results of a survey among participants in a large (n>10,000), longitudinal health study, and an Amazon Mechanical Turk population. Through these results, I locate brainscanning, and EEG specifically, as a fruitful case for understanding how particular sensing technologies surface and construct notions of mind.

Having motivated EEG as a fruitful sensing modality for further exploration, Chapter 6 shifts in focus from users to software engineers, studying their interactions with a working brain-based authentication system. This population's beliefs are of particular interest as consumer brainscanning devices become less expensive, and increasingly open to tinkering via software. Although we find a diverse set of beliefs among our participants, we discover a shared understanding of the mind as a physical entity that can and will be "read" by machines.

To conclude, chapter 7 proposes the term *telepathy* to describe the encoding and transmission of minds. I attempt to chart a path for future work, highlighting tensions between opportunities for novel computer-mediated communication, and concerns around security, privacy and surveillance. Finally, I propose telepathy as a way to understand not just what computers can know about the mind, but how machines may shape our notions of what minds *are*, and who we are as mind-having beings.

Chapter 2

Ants, Fungus & Telepathy

Would you wear a device in the workplace if your manager thought it could track your productivity, or creativity [15]? Would you allow your child to wear the same device in schools, where it could monitor both their academic achievement and their mental health [17]? Would you wear a fitness tracker if your resting heartrate could predict your future involvement in violent crime [62]?

In all of these examples, sensing technologies blur the line between *sensing bodies* and *sensing minds.* Today, increasingly inexpensive sensors with developer-friendly SDKs and APIs allow those with requisite software expertise to (purport to) detect phenomena ranging from mental health to mood, all without direct data about the brain [38].

In this chapter, I seek to dethrone the assumption that brain-scanning is necessary for computers to "read" or "decode" the mind. Drawing from contemporary theories of embodied, extended and distributed cognition, I argue that consumer sensing devices are already able to grasp at the contents of our minds by sensing our bodies, tools, and built environment (Section 2.1). I relate this argument to existing work in affective computing and computational social science, reframing them as having already begun the work of theorizing and building computational *models of minds* (Section 2.2).

Drawing on critiques of affective computing and computational social science, I center the primacy of human interpretation in both constructing models of minds, and interpreting the relevance of these models in the course of life. I propose this interpretive process as a starting point for understanding how models of minds might operate in the world (Section 2.3). I conclude by considering the limits of what computers can know about the human mind, and how beliefs about the mind structure these limits (Section 2).

2.1   Background

Consider the ant. The fungal complex *Ophiocordyceps unilateralis sensu lato* overtakes the ant's behavior without acting on its brain at all. Instead, it uses the ant's body to navigate the world, constructing a network of coordinated sensing and actuation atop the ant's muscles

[39]. By sensing the ant's environment and stimulating its muscles in response, it causes the ant to crawl beneath a twig and bite into it; once affixed to the twig, the fungus paralyzes the ant, using its body as a breeding ground (Figure 2.1).

Figure 2.1: *Ophiocordyceps unilateralis sensu lato* takes control of an ant's mind without input from its brain. By constructing a network of sensors and actuators atop its muscles, the fungal complex forces the ant to chew on the underside of a twig, after which the ant's body will serve only as a medium for fungal reproduction.

Ignoring questions of control, consider the degree of *sensing* the fungus must perform in order to utilize the ant's body. Using the ant's bodily infrastructure, the fungus creates a *model* of ant-experience robust enough to control the organism completely. Although the *Ophiocordyceps* fungal complex cannot read the ant's brain (it has no physical presence there), it can read the ant's *mind* well enough to model its environment and body. The fungus' model of ant-experience may not be the same, or even similar, to those used by the host ant. Regardless, they are of a sufficient resolution to allow the fungus to achieve its (reproductive) goals.

With this fungus in mind, consider the emerging class of internet of things (IoT) devices, which are increasingly embedded in the built environment, worn on the body, or worn inside the body via ingestible pills (Figure 2.2). Though common, cameras too sense bodies, often in public and without subjects' knowledge [90]. All of these connected devices are endowed to some degree with the capacity to sense (and to build models of) human bodies in space. Past work has referred to this process broadly as *biosensing*, and these devices as *biosensors* [30].

While humans are significantly more complex than ants, the *Ophiocordyceps* fungal complex helps illustrate the possibility of creating *models of minds* with limited or no

Figure 2.2: On the left, fungal filaments surround an ant's mandible muscle [39]. On the right, commercial sensing devices decorate the wrists of an enthusiastic self-tracker [33].

information from the brain. If fungus can do so, perhaps consumer sensing devices can, as well. As I review in this section, contemporary philosophical theories engage seriously with the notion of a beyond-the-brain mind. As I discuss in Section 2.2, these theories allow the physical phenomena detected by commercial sensors to be constituent of the mind.

Material theories of mind

What is the mind? What is its relationship to the body, and to the physical world? Philosophers have proposed two basic categories for answers to this question. *Dualism* posits that the mind has non-physical components, whereas *physicalism* posits a mind of only physical components (for a slice of this debate, see [20]). Since the biosensing aparatus I discuss here are restricted to physical phenomena, the dualist perspective presents an impasse for our analysis (how can physical sensors sense the non-physical)? The physicalist interpretation, on the other hand, lends itself naturally to scientific study—and to sensing. From the physicalist perspective, all phenomena in the mind can be reduced to descriptions of physical activity; thus, some physical theory will eventually explain the mind in entirety. The physical perspective provides a natural route forward for our analysis, as it implies that a sufficiently sensed physical world, combined with sufficiently robust theories about the mind, could yield a computational model of the mind.

The remainder of this section outlines various physicalist theories of the mind. Beginning with cognitive science's computational accounts of the mind, I trace critiques of this field to the newer theories of mind that have come to meet them. These theories motivate notions of beyond-the-brain mind, which in turn motivate the discussion on biosensors that follows in Section 2.2.

Cognitive science

Cognitive science has historically been an influential source of physicalist theories about the mind. The field takes a computational account of the brain, understanding how it "processes information" [103] within the physical constraints of computational space and time [92]. This perspective offers computational *models* of "cognition" [92]. For example, these models informed the design of neural networks, before the relatively recent discovery of performant backpropagation algorithms made neural networks practical to deploy [73].

However, cognitive scientific models of the mind have received considerable criticism [77, 103]. Two relevant critiques focus on cognitive science's "isolationist assumptions": a focus on the brain (isolated from the body), and a focus on the individual (isolated from social context, and from the environment). The following sections review major responses to these critiques: embodied cognition, distributed cognition, and extended cognition. These theories return later as I discuss prior work in affective computing and computational social science.

Mind extending into body: Embodied cognition

Cognitive science's isolation of the brain rests on the belief that the brain is strictly equivalent to the mind. This assumption has encountered two primary critiques. First, the dichotomy between the brain and body is unstable; neurons occur body-wide, running directly to the brain, such that it is difficult to evaluate the role of cerebral neural activity in the functions of mind irrespective of non-cerebral neural activity. Second, to quote Noë and Thompson (2004), "The exact way organisms are embodied simultaneously constrains and prescribes certain interactions within the environment." [77]. In other words, mind is manifested as it is due to the physical conditions of the body.

These critiques gave rise to the *Embodiment thesis*: that an agent's beyond-the-brain body plays a causal role in that agent's cognitive processing. For example, Noë and O'Regan's analysis of vision recasts the "visual processing" of cognitive science, in which internal representations are built and manipulated within the brain, to an active, embodied process, in which the world is not simply waiting to be seen, but actively providing its own representations; the body and brain must meet through an active process of co-adaptation [80].

Mind extending beyond body: Extended and distributed cognition

While the embodiment thesis prods at the causal relationship between mind and the physical conditions of the body, it glosses over the relationship between these bodies and the world in which they are situated. In response, Clark and Chalmer's *extended cognition* thesis argues that the environment at large can be considered as part of the mind; that "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" [26].

This theory does not stop at tools in describing a mind beyond the body. Broadly, extended cognition refocuses the brain away from the individual body, and toward the "active role of the environment in shaping cognition" [26]. This theory paved the way toward a socially-extended cognition, or *distributed cognition*, as described in Hutchins' (1995) ethnography of sailors on a naval vessel [51]. In his analysis, multiple individuals, and the material environment play constituent roles in cognition, manifesting a mind that is distributed across multiple human and non-human actors.

In addressing some critiques levied against cognitive science, the theories in this section make various cases for a mind that extends beyond the confines of the brain, and even beyond the confines of the body. The following section argues these theories, perhaps unwittingly, make the mind amenable to modeling via sensors that are worn or embedded in the environment, and that past research has (also unwittingly) already begun to sense the mind from beyond the brain.

2.2 Models of minds

The theories outlined in the previous section all propose that the mind is physically instantiated in the material world. They differ only in *where* this mind is said to exist, and where it does its work. Using these theories, this section argues that two prior research programs have already attempted to sense aspects of mind from beyond-the-brain bodies.

To assist in this analysis, I propose term *models of minds*. This term borrows from autism research's *theory of mind*, which refers to the (human) ability to reason about mental states [7]. By substituting the word "theory" with the word "model," I emphasize formal or algorithmic representations. By then turning this singular "model of mind" into a plural *models of minds*, I highlight the intrinsic contestability of the algorithms that build them, the beliefs that underlie their construction, and the diversity of minds in the world to model. The term aims to cast a subtle doubt on models that appear too simple, or which (cl)aim to generalize too broadly.

In the remainder of this section, I read two strands of existing work through different accounts of mind: affective computing through embodied cognition, and computational social science through distributed and embodied cognition. I argue that physical theories of the mind allow these two fields to claim that they sense the ground truth of mental phenomena. Thus, I argue that these fields have already begun the work of building models of minds using data from the beyond-the-brain-body.

Affective computing

Affective computing, pioneered by Rosalind Picard at the MIT Media Lab, seeks to use sensors to measure a users' affect, emotions, and mood in order to improve their interaction with machines. [83]. Two commercial examples of such sensing come directly from work in Rosalind Picard's research group. The Empatica wristband senses electrodermal activity,

with the aim of correlating these data to emotional states [41]. This wristband has gone on to inspire cheaper consumer alternatives, such as the Feel [38]. Also from Picard's lab, Affectiva classifies emotions from facial expressions, as detected through a camera. Their infrastructure works through a webcam, providing what they term "Emotion as a Service" [1].

In both of these examples, affect is framed as a bodily state, as in theories of embodied cognition. However, affective computing extends these claims further, positing that wearable sensors can measure, encode, and transmit emotions through their sensing of bodily states [49]. Although work in affective computing does not generally make explicit references to embodied cognition, it typically seeks to detect emotion via bodily phenomena, and does not consider these phenomena to be proxies from real emotions, indicating a general view of emotions as embodied primarily.

Computational social science

In this section, I argue that distributed and extended cognition allow past work in computational social science to claim that these sensors can detect the ground truth of mental phenomena. Past work in computational social science has used mobile sensors as sources of data about human interaction, efforts that predate both commercial IoT devices and the general ubiquity of smartphones in the global north. One early example is Sandy Pentland's sociometer, an internet-connected necklace outfitted with a variety of sensors [79]. In contrast to Picard's affective measurements from single users, Pentland's work measures phenomena distributed across multiple individuals.

The Social fMRI provides a seminal example. A distributed, multimodal sensing infrastructure, implemented via mobile phones over more than a year, aimed at sensing "how things spread in [a] community, such as ideas, decisions, mood, or the seasonal flu" [3]. In this frame, both "ideas" and "the flu" are equated as properties not of individuals, but of communities and relationships. The Social fMRI study spawned numerous, similar projects, including one explicitly aimed at detecting "happiness" [12] or "creativity" [15], and, relevant to our discussion, one that aimed to diagnose depression from mobile phone traces [17]. In this study, longitudinal GPS traces were correlated with answers on questionnaires via machine learning and related statistical techniques.

As embodied cognition allows affective computing to present bodily phenomena as constituent of emotions, distributed and extended cognition allow this work to present extrabodily and multi-individual phenomena as constituent of mental states. If one believes depression to be an embodied phenomenon then the phone could be said to sense depression's bodily correlates. However, if one believes depression to be an extended phenomenon, then the cellphone could in fact be *a constituent of the depression itself*, to report the ground truth of depression. Distributed and extended cognition are instructive in understanding how technical artifacts might seek the ground truth of phenomena relating to the mind, such that models can be said to be accurate or inaccurate.

In the next section, I review critiques of the work discussed above. I use these critiques to center the role of human interpretation in building models of minds and in making them

legible in social context.

2.3 Centrality of interpretation

Today, the world of computational social science has informed the commercial world of targeted advertisements; affective computing has begun to creep into our lived experience, with consumer devices that purport to continuously measure emotions [38].

However, their legacies live on. Computational social science, for example, relied heavily on top-down maps (the Social fMRI paper included a figure with an eye looking downward [3]). This top-down purview of the scientist eschewed potential concerns around individual privacy, a legacy that continues to produce struggles in IoT. Consider the contemporary example of Uber's employees-only "god view," which makes visible the location and movements of all users and drivers [78, 46]. The persistence of top-down perspective in modern work gestures broadly to the ways beliefs and assumptions can be fed forward from academic studies into commercial products, becoming ensconced in technical artifacts.

Given the ongoing relevance of affective computing and computational social science in our emerging world of pervasive biosensing, this section reviews some of the most pointed critiques these fields have encountered. These critiques center the role of human interpretation in making models of minds buildable (by engineers) and legible (to end-users) in social context. In supporting this perspective, I review past work on how people bring signals from the body to bear on the mind.

Attempts to classify or detect mental phenomena have faced a variety of critiques. First, these studies have tended to frame mental states as definite entities for which a single ground truth exists. Boehner et al [10] propose an alternative: emotions as co-constructed, performed socially, and understood only in collaboration with other socially-experiencing subjects. An account of socially situated emotions has received some limited uptake within affective computing [82]. However, these theories still pre-categorize emotions, obscuring phenomena at the borders of these categories [10]. This critique effectively posits that beliefs about the mind limit what phenomena can be modeled or sensed. The mere invention of categories precludes detecting phenomena outside of their borders, and may even preclude finding phenomena that lies between categories.

Second, little work yet has substantively engaged with the question of how algorithms and devices that seek to detect emotion may affect the way emotion is experienced or performed. Past work strongly indicates that feedback about emotional experience may alter the way emotions are experienced [95], and that context may radically alter the way these models are understood [69]. In this critique, beliefs about the mind strongly inform and structure what can be *understood* about the mind from a given model.

Bringing signals from the body to bear on the mind

If beliefs about the mind structure interpretations of biosensory data, then how do these interpretations about the mind come to be? I argue that the meaning of biosignals are shaped by prior beliefs about the body, as well particular social contexts [69, 4, 95]. Through past work, I outline how the suggestion provided by particular sensing devices can meet with pre-existing beliefs about the body, producing socially-relevant interpretations regarding the mind.

In Ali et al (2014), undergraduates in neuroscience believed a "scanner" (in reality, a perm machine from a hair salon, painted gray) could read their thoughts in some detail, even after the researchers told them explicitly that such technology is not (yet) possible [4]. The authors suggest that this indicates people have some intrinsic faith in brainscanning, perhaps due to "neurohype" in popular media [99]. Another way of interpreting this finding, however, is that *biosensing systems* offer a particular white lab-coat effect of their own, which interacts with social context to produce specific interpretations. This latter proposal is suggested by [95], in which the *Moodlight* is able to make people feel relaxed, simply by suggesting that the person is relaxed already. From the user's perspective, either that the machine "knows better" than they do, or that people fill in the gaps in their ability to introspect using the machine's determination. This interpretation is also suggested by [9], in which the amount of time people were talking in a group conversation was displayed visually on a table. This study finds that people are willing to believe some distortion, but only to a point. Interfaces provide suggestions, which end users may accept even when they conflict with what users feel to be true.

However, suggestibility does not entirely account for why people build interpretations about the mind from sensor data. People bring beliefs to the table as well, which structure what they are willing to accept. For example, the results observed in [4] had something to do with the fact that the machine was scanning the brain; if it had been taking a saliva sample, for example, subjects may not have been as likely to believe it could detect their thoughts. In other words, beliefs about what biosensing devices can capture about the mind are a product both of particular interfaces, and their pre-existing beliefs about the body, and the relationship between the body and phenomena in the mind. These beliefs may vary with culture, as well. We have no particular reason to think they are any more universal than, e.g. the perception of color [86].

The central role that beliefs about the body play are reinforced by studies on ubiquitous heartrate sharing. Heartrate sensors have been among the first physiological sensors to be widely embedded in consumer devices, usually in smartwatches or earbuds. Slovák (2012)'s study on heartrate sharing [93] revealed that beliefs about heartrate can take on meanings that relate intrinsically to the presumed meaning of hearts and heartrate. In [69], we found that an elevated heartrate signal correlated with reduced cooperation in an iterated trust game, where elevated "SRI" (a fictitious biosignal) did not. These studies indicate that beliefs about the body, originating either from media, or embodied experience, have some effect in suggesting possible meanings for biosignals in social context.

How minds are made and modeled

The case of affective computing in relation to embodied cognition, or of computational social science in relation to distributed and extended cognition, illustrate how beliefs about the mind inform, shape and structure the claims that technical practitioners make about the artifacts they design. Although these projects did not explicitly cite philosophical progenitors, their shared perspectives on the mind afforded their success in detecting phenomena such as emotion or mental health.

Given the lasting impact of not just these research programs, but the perspectives they embed, it is critical to review the perspectives of these programs and their antecedents. How do these academic disciplines inform technical practice on the ground, particularly among software engieners? The perspectives of engineers are relevant to understanding what they build, and why. Some past work has looked at engineers beliefs with respect to sensing devices. For example, Sample's work on neuroengineers [88] and my own work on software engineers [70] have examined engineers' complex and heterogeneous beliefs about the mind and body.

In tandem with the beliefs of engineers, users' beliefs about the mind, formal or informal, also inform, shape and structure what users believe, or are willing to believe. To quote Dawn Nafus as she described her early studies in biosensing, "figuring out whether a consumer market for biosensors was even thinkable had everything to do with whether the data they produced cohered with a cultural and social imaginary, such that users stood a chance of making sense of them" [74].

In this chapter, I reviewed how beliefs in theories about the mind (formal or informal) play a critical role in defining how models of minds are built, and how they are understood as relevant in context. While we will return to the question of how models of minds are built in Chapter 6, the following chapter will look at how end-users interpret models of minds in social context. The two studies described there will demonstrate how people use basic biosignals in computer-mediated contexts to build interpretations relating to the minds of others.

Chapter 3

Reading mind from heartrate

The previous chapter argues that human interpretations are central to the study of how models of minds might operate in the course of life. Building on this argument, the present chapter seeks to uncover what users believe basic biosensors can capture about the minds of others. Through a vignette experiment and a mixed-methods experimental study, this chapter show how people use biosensory data (heartrate) in social, computer-mediated contexts to build interpretations relating to the minds of others.

3.1   Background

As of 2016, several apps allow users to share their heartrate with their friends, leading some [67] to wonder why anyone would anyone want to do such a thing. In fact, heartrate is a potentially rich signal for designers. The meaning of a heartrate in any given context is at once socially informative [40, 93] and highly ambiguous [68].

After all, heartrate is not just some number. The sense of one's heartbeat is an integral feature of the human experience, and people's associations with it range from intimacy [56] to anxiety [31] to sexual arousal [100]. Many heartrate sharing applications rely on these associations, asking users to ascribe contextual meanings to heartrate [57, 93], often with the aim of increasing intimacy [56]. 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? [18]

These applications, along with many others, rely on the fact that people will imbue their heartrate data with emotional, and highly contextual interpretations. Given the relatively large number of wearables with embedded heartrate monitors (watches, bands, even earbuds) [97], it is unsurprising that designers are looking beyond fitness and health for ways to increase

user engagement with these devices. However, it is not clear how individuals will interpret a shared biosignal (e.g., heartrate) in different contexts of social interaction.

This chapter examines what heartrate can mean as a computer-mediated cue, and how interpretations of heartrate affect social attitudes and social behavior as people assign meanings to these signals relevant to the mind (emotion, mood, trust).

First, we use a vignette experiment to 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 [22]. 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), in which the person they are meeting is running late. I discuss the vingnette in more detail later.

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.

Sharing sensor data

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 [42] (p 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 [50]. 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.

Sharing physiological data

Prior work interrogates the contextual interpretation of personal data from certain kinds of sensors [23, 27], 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 represents a one-to-one mapping to a point on the surface of a sphere, heartrates do not have one-to-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 computer-mediated 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 (e.g., [9], [59]). Another class of prototypes explores how signals might affect feelings of intimacy, particularly between romantic partners [8], and several applications focus on the transmission of heartrate as a means to achieve this effect [56, 67].

Sharing heartrate

Heartrate has deep-rooted cultural significance in many societies, and near-universal familiarity as a feature of our lived experiences. Building on associations with intimacy and love, many heartrate sharing applications have aimed to "enhance" social connectedness by fostering feelings of intimacy between people [56, 47].

What heartrate means as a computer-mediated cue, however, is ambiguous, its potential interpretations varying widely in different contexts [66, 93]. Boehner et al (2007) argue for the intrinsic ambiguity of sensor data as a resource in design, particularly in systems that seek to use these data to express emotion [10]. Many technology probes corroborate this stance, relying on users to project socially contextual meanings around a transmitted heartrate. Consequently, more recent work has challenged the notion that the social consequences of transmitting physiological data will always result in increased trust and intimacy [93]. There remains little work, however, on how the potential ambiguity of a heartrate signal is resolved in social conditions of risk and uncertainty.

3.2 Vignette experiment

This section describes the quantitative and qualitative results of a randomized vignette experiment in which subjects (103 undergraduate students) 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.

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 [81] 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 [106].

If lay interpretations of one's own heartrate can yield negative self-interpretations [81, 106], sharing heartrate information could also yield negative social interpretations of 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.

Hypotheses

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.

Past work indicates that people tend to make negative inferences about mood and emotion from elevated heartrates [31, 45, 106]. 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 less calm (1a), more emotional (1b), and more easily upset (1c), compared to those who believe that their partner has a normal heartrate.

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.

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.

Sample

Our sample consisted of undergraduate students recruited from the population of UC Berkeley. Potential participants were asked to participate in a short online survey; they did not know the nature of the questions or the topic of the study in advance. All the participants were compensated with a $5 Amazon gift card. One hundred and three (103) participants completed the experiment survey instrument. The pool was weighted toward women: 65% were women and 34% were 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.

3.3 Quantitative results
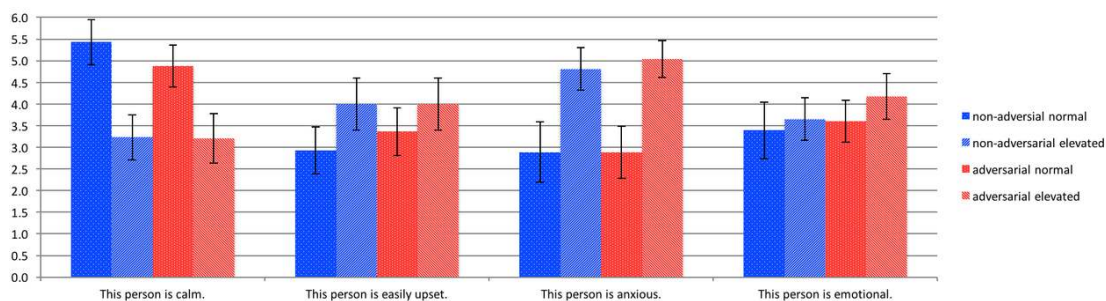
Figure 3.1: Mood-related evaluation (7-point Likert) means by condition (bars represent standard deviation).

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
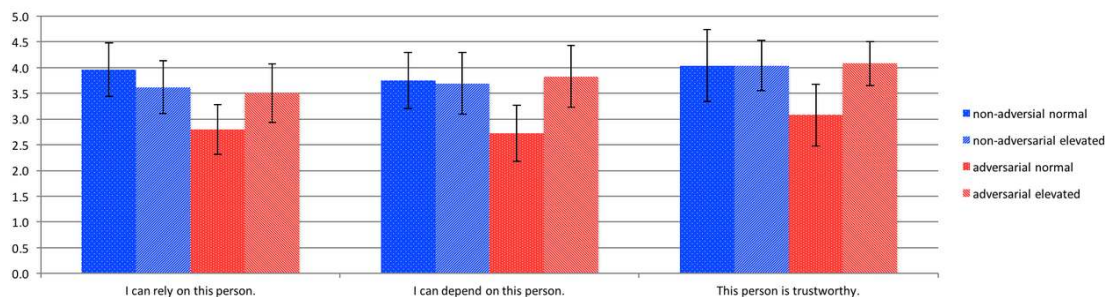
Figure 3.2: Trust-related evaluation means (7-point Likert) by condition (bars represent standard deviation).

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 3.1). 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, $F(3, 97) = 4.19$, $p < .01$; 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 3.2). 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.

3.4   Qualitative results

Directly after the vignette, participants were asked four free-response 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.

Adversarial context / Normal heartrate

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.e., stressed that s/he is running late).

> 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.

Adversarial context / 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. We return to this point in the discussion.

Non-adversarial / Normal heartrate

In the non-adversarial context (meeting for a movie), 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. 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.

Non-adversarial / 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.

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.

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.

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.

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.

3.5   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 [106]. 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 trust-related 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. Computer-mediated communication researchers have long noted that, when cues are omitted from computer-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 people's beliefs about what machines can know about the mind. 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. The social situatedness of models of minds are probed further in this dissertation, particularly in chapters 4 and 6.

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.

In our qualitative data, we regularly observed attitudes about the presumed authority or "neutrality" of data interacting with beliefs about the body to create 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. This point motivates the work described in Chapter 5, which broadens this inquiry to a variety of sensors and a variety of aspects of mind.

3.6 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 influence social interpretation. 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.

While this vignette experiment provides evidence that interpretations of biosignals from sensors (such as wearables) can affect social attributions and behaviors towards others. Nevertheless, many questions remain. While this study examined social beliefs as they relate to heartrate, it did not examine how (or if) these beliefs affect social behaviors. Furthermore, we did not examine how specific our findings are to heartrate. What other signals from the body might lead to social interpretations?

3.7 Conclusion

In the following chapter, we begin to address the limitations above through controlled, behavioral experiments, which help us ask more specific questions about how elevated heartrate affects perceptions of risk in uncertain interactions, e.g., when money is at stake. This study leads to a more robust understanding of how the transmission of basic biosignals might affect social behavior.

Chapter 4

Biosignals, mind and behavior

From the prior chapter's findings about social attitudes, this chapter moves to a lab-based experiment to understand how shared heartrate effects social behavior. We apply quantitative and qualitative analyses to an iterated prisoner's dilemma game, in which heartrate information ("elevated" or "normal") was shared between players. In a follow-up study, we replicate our initial study, but replace heartrate with an unfamiliar biosignal, "Skin Reflectivity Index (SRI)."

We find that both heartrate and the unfamiliar biosignal, when elevated, are associated with negative mood attributions, but we observe a decrease in cooperative behavior only with elevated heartrate. Qualitative results indicate that individuals may learn an association between our unfamiliar biosignal and the cooperative, trusting behavior of their partner. Our findings highlight the role prior beliefs can play in shaping interpretations of a biosignal, while suggesting that, in the absense of prior beliefs about a particular signal, users may learn to associate signals with social meanings over repated interactions.

Our results raise important questions for applications that transmit sensor-derived signals socially between users. For signals with strong cultural associations, people's prior beliefs will color their interpretations, and social outcomes may or may not be positive. In the case of novel signals, on the other hand, our results imply that designers can (perhaps inadvertently) teach users to associate these biosignals with social meanings. This effect could be viewed as beneficial, depending on design objectives. It could also be dangerous if designers suggest, perhaps even inadvertently, interpretations that lead to discrimination.

4.1 Lab-based experiment

Following our vignette experiment in the previous chapter, which focused on social attitudes, we extend our inquiry to a trust-building game, which will allow us to study social behavior. Through quantitative and qualitative analyses, we find that "elevated" (versus "normal") heartrate of an exchange partner is associated with negative mood attributions and reduced cooperation in a social dilemma game. To investigate how specific our findings are to

heartrate (as opposed to some other "elevated" signal collected from the body), we replicate our initial experiment with a fictitious biosignal, "skin reflectivity," which will be unfamiliar to participants. We find that both heartrate and the fictitious biosignal are associated with negative mood attributions, but we observe a decrease in cooperative behavior only with elevated heartrate. Qualitative results indicate that individuals may learn an association between our fictitious biosignal and the cooperative, trusting behavior of their partner. Our findings highlight the role prior beliefs can play in shaping interpretations of a biosignal, while suggesting that designers can, perhaps inadvertently, train users to associate signals with social meanings. We discuss implications for how wearable sensors can mediate social interactions.

Generally when individuals believe that their heartrate is elevated, they often believe their mood and emotions to be more negative [100]. Thus, we apply this same logic to how individuals will interpret the elevated heartrates of others in uncertain social interactions:

> **Hypothesis 1**: Participants who see a consistently elevated heartrate from their partner will rate their partner more negatively on mood attributes, compared to participants who see a consistently normal heartrate in uncertain and risky social interactions.

If elevated heartrate has a negative connotation with mood, then elevated heartrate may increase uncertainty about the behavior of one's partner as well. When people know that their partner has an elevated heartrate in an uncertain, risky interactions, they may take actions to protect themselves against potential losses. In trust-building situations, individuals take small risks with other people (entrustment behavior) and learn whether the other person honors that trust or not (cooperative behavior). Thus, individuals have two different ways to respond to increased uncertainty about their partners' behavior in trust situations: 1) reduce the amount they entrust to their partners, or 2) decrease their willingness to cooperate with the partner [22, 28]. Since we expect elevated heartrate to have pre-existing connotations with negative attributes, we predict that individuals will entrust and/or cooperate less to protect themselves from potential harm when the partner has an elevated vs. a normal heartrate.

> **Hypothesis 2**: Participants who see an elevated heartrate from their partner will (2a) trust less, and (2b) cooperate less with the partner in uncertain and risky social interactions compared to participants who see a normal heartrate.

4.2   Sharing heartrate in a risky, uncertain interaction

In order to test our hypotheses, we conducted a repeated trust experiment with shared heartrate information. Trust games present participants with financial incentives to pay attention to their partner's decisions over time, and provide means for operationalizing trust and cooperation in the presence of uncertainty [22].

The overall design of the trust game involves anonymous pairs of fixed partners making repeated decisions to entrust valued resources to the partner, and to return (cooperate) or keep (defect) the points entrusted by the other partner. Importantly, individuals can make the highest amount of money when they entrust many points to a partner and the partner returns these points. This creates an uncertain social situation in which participants are trying to earn real money by repeatedly taking risks (entrusting points) to a partner. Since the partners are making the same decisions to entrust and keep/return points from the other partner, these are mutually-dependent social interactions.

Experimental Design and Methods
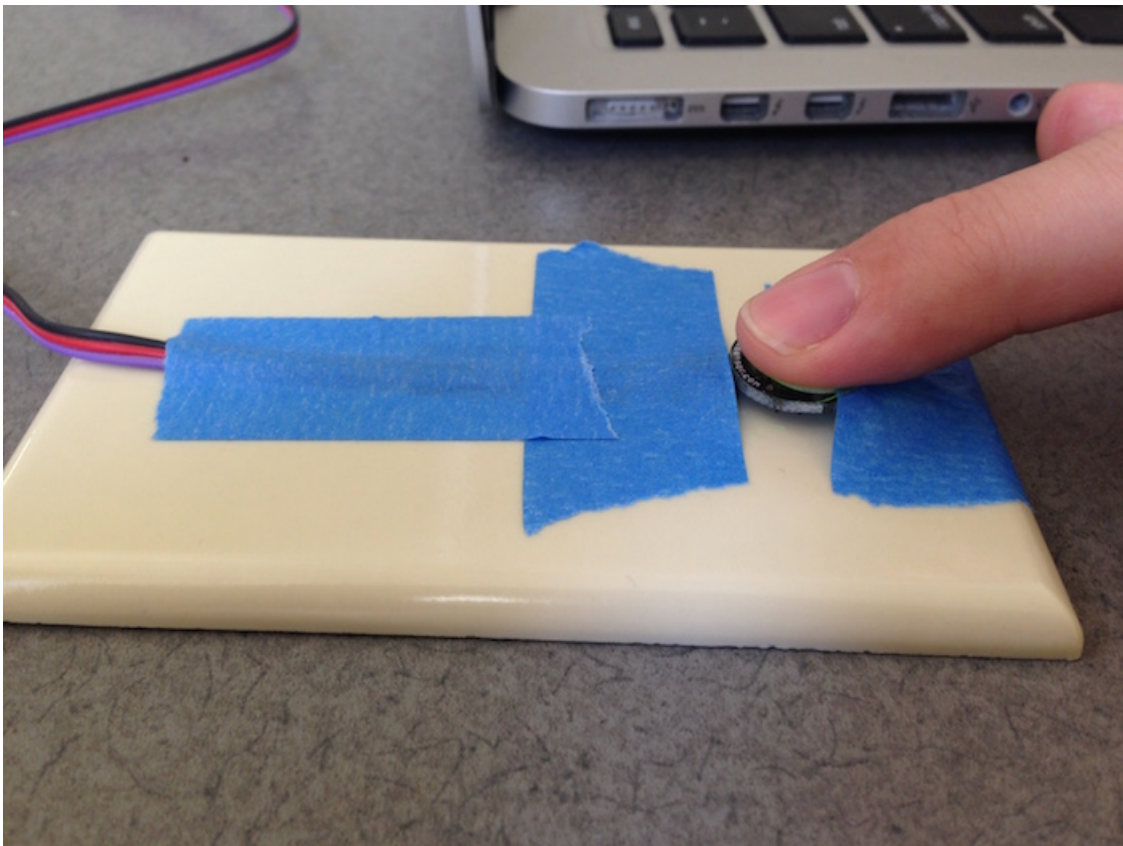
Figure 4.1: The heartrate monitor. Participants were told to place their finger on the monitor to take a reading while viewing their partner's decisions during the previous turn.

We operationalized an uncertain social interaction situation using a trust game called the Prisoner's Dilemma with Dependence (PDD) [22, 28]. The PDD game allows individuals to control the amount of risk that they want to take with their partner by choosing how many points to entrust, followed by a second decision to either keep or return whatever has been

entrusted by their partner. Thus, the PDD game separates trust behavior (choosing how much to entrust to a partner) from cooperative behavior (choosing to return or keep what a partner entrusted). In each round of the PDD game, participants were given an initial endowment of 10 points. Each participant decided whether to entrust any number of points to their partner, from zero to ten. Then, participants found out at the same time whether their partner had entrusted them with any of their own points, and if so, how many. Next, each participant decided whether to keep the points entrusted to them (defection) or return them (cooperation). The participants could not return only a portion of the entrusted points, only all or none of them. If the points were returned to the partner, they were automatically doubled in value for that participant.

After all participants made decisions about returning or keeping any points that had been entrusted to them, they were then asked to place their finger on the heartrate monitor for a few seconds in order to get a pulse reading (Figure 4.1). Participants then viewed the summary of point calculations for the round. Subsequently, participants viewed a visual display of the partners' recent heartrate (Figure 4.2). The final point calculation for the round included any of the initial allotment of points remaining after the trust decision, plus any points that the participant kept from their partner if they decided not to return them. In addition, players received points for any entrusted points that their partner returned, which doubled in value.

When participants arrived at the laboratory, they were given a consent form that described the nature of the study, as well as the human subjects' approval information from our university. We wanted participants to believe that they would be interacting with other real people, and this perception was enhanced by having 12-16 participants at separate computer terminals in the same large room during each experimental session. In fact, we controlled the trust and cooperation behavior of the "partner" for every participant using a simulated computer actor. As a result, no one in the study interacted with a human partner.

The simulated actor was programmed to always begin by entrusting one point on the first round, then randomly entrust up to one point above or below whatever the partner entrusted on the previous round. In addition, the simulated actor was programmed to always cooperate (i.e., return the points that were entrusted by the partner). Following [22], we chose to use a highly cooperative interaction partner in order to minimize any other forms of uncertainty in the interaction. A highly-cooperation partner does not introduce any defection behaviors that might otherwise reduce cooperation or trust from the participant (thereby hindering our ability to detect main effects from the experimental manipulation). Thus, the simulated actor was designed to reciprocate the entrusting behavior of the human participant on each round, and always cooperate no matter what the human participant chose to do.

The participants completed 20 rounds of the PDD game, but they did not know how many rounds they would play in order to eliminate end-game effects, such as defecting at the last minute. After all rounds of the PDD game were completed, participants answered a short post-questionnaire in order to assess their attitudes and beliefs about their partner. This questionnaire included 7- point Likert-style response questions (1 = strongly disagree, 7 = strongly agree) about the partners' beliefs about the partners' anxiety (e.g., "my partner is

anxious" and "my partner is calm").

As a manipulation check on the perceptions of the simulated actor's behavior, we also asked questions about the partners' game behavior ("my partner is trustworthy" and "my partner is cooperative"). Finally, we supplemented our quantitative measures with two open-ended questions: "How would you describe your partner?" and "What, if anything, did heartrate tell you about your partner during this experiment?" Participants were paid between $15-30 based on their point earnings during the game. The entire study lasted one hour.
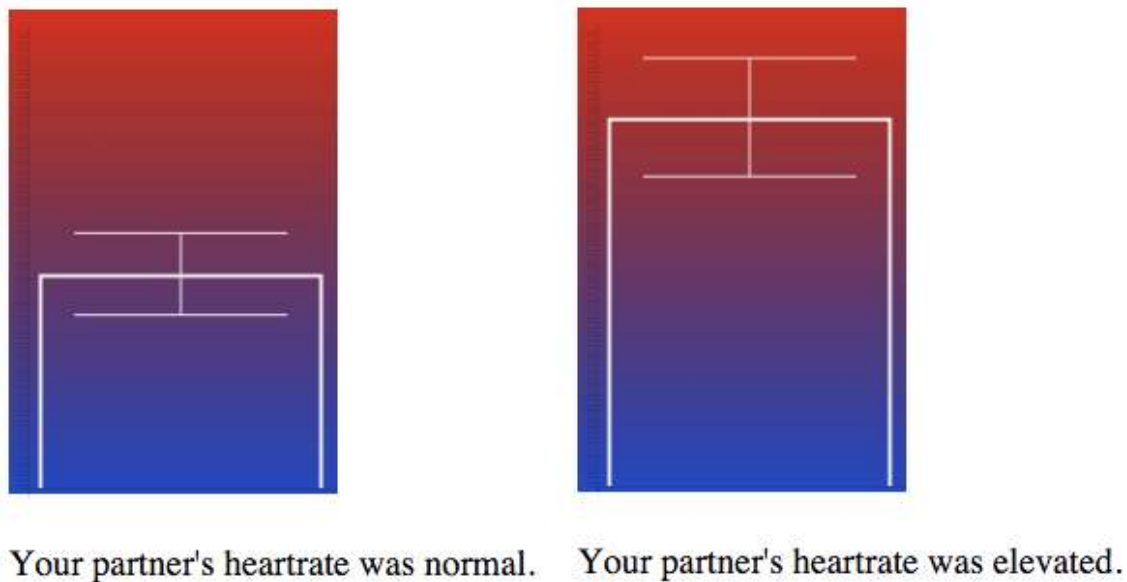
Figure 4.2: The heartrate visualization. After viewing the results of the previous round, participants saw a graph of what they believed to be their partner's heartrate, either normal (left) or elevated (right). Error bars fluctuated within pre-set bounds.

At the end of the study, participants were debriefed on the true nature and intent of the experiment. An experimenter was available at the end of the study in case of any questions, and we provided participants with the researchers' email addresses on both the signed informed consent form, as well as the debrief form, so that they could contact us regarding any aspect of the study. We did not receive any emails or concerns from participants.

Experimental Manipulation

To assess the effect of interacting with a partner who has an elevated heartrate versus interacting with a partner who has a normal heartrate, we controlled the heartrate information that participants saw after each round of the experiment. This created a two-condition design: always normal heartrate (NH) and always elevated heartrate (EH).

Participants and Procedure

Our sample was undergraduate students recruited from the population of a large, public university on the West Coast of the United States. We contacted potential participants via email from a voluntary experimental subject pool. All participants expected to be contacted to participate in a social research study at some point during the semester, and knew that they would earn between $15-30 during this one-hour study, depending on their choices during the experiment. Fifty-six participants (56) completed the experiment, 41 women, 14 men, and one self-identified as other. The mean age of participants was 21.

Upon arrival at the laboratory, participants were guided to an individual desk with privacy walls. After signing an informed consent form, participants read written instructions on the computer which explained that they will have the opportunity to interact with a single partner for many rounds in order to examine decision making in social situations. Participants were also told that we would collect pulse (heart rate) information at designated times during the study using a simple pulse monitor that was connected to the laptop computer.

Validity Check of the Visualization

Our study aims to understand the effect of "elevated," as compared to "normal," heartrate. As such, we needed to show participants a visualization that afforded only a relative value for heartrate, not an exact figure (since different people may have different ideas of what number value constitutes a normal or elevated heartrate).

We designed a visualization to display a relative heartrate (Figure 4.2) and performed a small (n=25) face validity check to ensure that our visualization would work as intended in the actual experiment. In our short validity survey, we included three versions of the visualization, representing a mix of elevated, low and normal heartrate, and two Likertscale questions: "The precise meaning of this graphic is ambiguous," and "I can interpret the difference between 'low', 'normal', and 'high' heartrate from this graphic," which participants answered from "Strongly Agree" to "Strongly Disagree" on a 5-point scale. We also included two open-ended questions, "Please explain what the picture is telling you about one's heartrate," and "Please explain what this picture does not tell you about one's heartrate."

We distributed this survey over an email list to students and alumni of a public, West Coast US university, and received 25 valid responses. The answers to both Likert questions indicated agreement that the visualization was both ambiguous (mean = 3.58, S.D. = 1.28) and also easily interpretable (mean = 3.41, S.D. = 1.35). Importantly, open-ended qualitative responses confirmed that the heartrate was easily understandable, but that the precise value of heartrate was ambiguous.
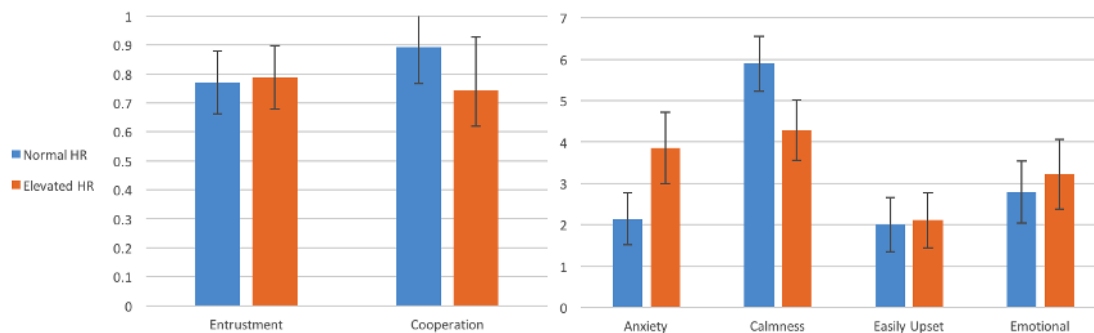
Figure 4.3: Means of entrustment and cooperation (left) and mood attributions (right) in elevated and normal heartrate conditions.

4.3 Results

Quantitative results

Our first hypothesis predicts that, when individuals believe that their partner has a consistently elevated heartrate, compared to a normal heartrate, they will rate the partner more negatively on mood attributes. Consistent with prior research, we found an overall strong, statistically significant effect and medium practical association between attributions and experimental condition, $F(4, 51) = 6.7$, $p < .0001$; Wilk's lambda $= .66$, partial eta squared $= .34$. Turning to the individual outcomes, we find that perceptions of the partners' anxiety is significantly higher in the EH condition ($M = 3.86$, $SD = 1.72$) compared to the NH condition ($M = 2.14$, $SD = 1.27$), $F(1, 54) = 18$, $p < .001$; partial eta squared $= .25$. Furthermore, participants rated their partners as significantly more calm in the NH condition ($M = 5.9$, $SD = 1.3$) compared to the EH condition ($M = 4.29$, $SD = 1.46$), $F(1, 54) = 18.71$ $p < .001$; partial eta squared $= .26$. On the other hand, we found no statistically significant differences for perception that the partner is "easily upset" or that the partner is "emotional" ($p = $ n.s.). In sum, we find strong statistical and practical differences in perceptions of both anxiety and calmness, but no statistical or practical differences in perceptions of how emotional or easily upset the partner is in the two experimental conditions. Given the significant omnibus test and significant results on two of the four individual outcomes, Hypothesis 1 is partially supported.

Our second set of hypotheses predict that participants in the elevated heartrate (EH) condition will exhibit lower trusting (H2a) and/or cooperative (H2b) behavior compared to those in the normal heartrate (NH) condition. The average points entrusted by participants in the EH condition ($M = 7.88$, $SD = 2.18$) was not significantly different than the NH condition ($M = 7.7$, $SD = 2.18$), $t = .28$, $p =$ n.s, one-tailed test. Thus, individuals entrusted points to their partners at approximately the same level in both conditions (Figure 4.3). Hypothesis 2a is not supported.

However, we found that the average cooperation rate in the EH condition ($M = .74$, SD

= .37) was statistically significantly lower than the NH condition (M = .89, SD = .25), t = 1.76, p < .05, one-tailed test. Importantly, this result shows a medium practical effect size (Cohen's d = .47), indicating a meaningful real world difference. On average, those in the normal heartrate condition cooperated 20% more than those in the elevated heartrate condition (Figure 4.3). Hypothesis 2b is supported.

Since we designed the simulated actors in both conditions with trusting and always-cooperative behavior, we did not expect participants to rate the simulated actors differently in terms of the focal behaviors of cooperativeness and trustworthiness between experimental conditions. This is a critical manipulation check, since we need to rule out any perceived effect of the simulated partners' behavior in order to establish that the primary treatment (heartrate of partner) had an effect on the human participants' behavior. The omnibus test of difference in perceptions of the trustworthiness and cooperative behavior between conditions was not significant, F(2, 53) = .21, p = n.s.; Wilk's lambda = .99, partial eta squared =.01. Thus, as we would expect, individuals did not indicate significant behavioral differences for the trusting, cooperative simulated actor (which was programmed to behave exactly the same in both conditions).

Qualitative results

At the end of our questionnaire, before the demographic questions and the debriefing, participants were presented with two open-ended questions. The first asked participants to "Tell us how you would describe your partner." The second asked participants "What, if anything, did heartrate tell you about your partner during this experiment?" This section discusses and unpacks some of the responses that these questions elicited.

Many people who referred to elevated heartrate in their responses mentioned that it signaled anxiety. In some cases, participants even reflected on a negative relationship between elevated heartrate, anxiety and trust:

> how excited he/she is, whether he/she cheated

> It was elevated all the time so I think s/he was anxious [...] so I guess s/he did not completely trust me

These quotes further support our first hypothesis, as well as findings of past work showing that elevated heartrate typically signals anxiety and mood. In other words, elevated heartrate (and heartrate in general) seemed to be about the partner's current disposition, rather than who the partner was as a person. While the majority of those who mentioned elevated heartrate implied a causal relationship between the signal and the game context, a few did not:

> My partner's heart rate was elevated the whole time, most students are stressed so that might be why.

> They may have been nervous because of doing the experiment itself.

The relative rarity of skepticism about the relationship between heartrate and specific game events highlights the crucial role of framing and salience in turning what might be a disembodied signal (heartrate data) into a relevant, contextual clue. We also noted diversity in beliefs about the meaning of heartrate itself. Where almost all participants who mentioned heartrate associated it with anxiety, at least one participant had an entirely different take on his/her partner's consistently elevated heartrate:

> My partner's heart rate does not change too much which indicates that he or she is very nice.

These quotes highlight overall diversity in what an elevated heartrate is capable of meaning. Even within our relatively small, and relatively homogenous sample of university students, our quotes imply a mostly negative association with elevated heartrate, but also a potentially long tail of diverse beliefs about elevated heartrate.

Many participants said that normal heartrate indicated that the partner was "calm," "chilled out,' or "not anxious."

> [HR signaled] that my partner was always calm. The heart rate never fluctuated, it didn't make a difference.

> They remained calm

> I think it showed that my partner wasn't too nervous to see if he/she was returned the points or not, maybe because it was just an experiment or maybe because he/she wasn't worried about what result he/she was about to see was.

These quotes show subjects inferring a direct connection between the heartrate signal and the attribution of a calm mood. One participant specifically mentioned that consistency of normal heartrate made their partner seem more trustworthy:

> My partner's heart rate has been consistently normal throughout the experiment, so I guess s/he has no intention to cheat.

Another participant, presumably a cooperative one, thought that their partner's heartrate would have risen if s/he had not cooperated:

> I think it remained the same [normal] because I paralleled my partner's actions whereas if I had contradicted them, their heartrate probably would have changed in response.

In all of the above quotes (and the vast majority of responses), participants inferred a relationship between normal heartrate and calmness. However, a few participants did not infer any relationships between behavior, moods and the signal they saw.

> Heartrate did not tell me anything. My partner was average each time. I also am sure I have an elevated heart rate due to coffee consumption so I did not take my partners into consideration.

> I based my decisions on their previous actions.

Not every participant explicitly inferred a calm mood from the normal heartrate signal, but most did. Taken alongside our quantitative results, our qualitative results provide evidence that subjects have used the emotional attributions they made based on their partner's normal heartrate to guide their behavior in the trust game.

4.4 Sharing an unknown signal in a risky, uncertain interaction

In the prior experiment, we found that participants cooperate less with partners who have elevated heartrates in the repeated trust game, compared to those with normal heartrates. While this result supports one of our key hypotheses, it also begs another question: Is the effect we observe due to heartrate specifically, or might any elevated biosignal show the same results for negative perceptions of mood and reduced cooperative behavior towards the partner?

In our second experiment, we attempt to tease out the effect of the heartrate signal itself, compared to any "elevated" (versus "normal") signal collected from the body. We replicate the first study, except that we tell participants that our monitor device measures SRI (Skin Reflectivity Index). SRI is an unfamiliar biosignal, for which individuals should not have any prior cultural or social beliefs.

Hypotheses

Without any context for what SRI means as a signal, participants may assume that any biological signal that is "elevated" from normal will be negatively associated with one's mood. If this is the case, then we should observe the same general pattern of negative mood attributions and less cooperative behavior when the partner has an elevated SRI as we observed with heartrate.

On the other hand, perhaps heartrate is special due to its common social associations with mood, anxiety, and even deception. If heartrate is distinctive in this regard, then we would not observe the same significant differences between normal and elevated SRI and mood attributes, trust, and cooperation rates with the partner.

To test the effect of our unfamiliar biosignal on behavior in risky, uncertain interactions, we evaluate the exact same hypotheses from study 1 again in the context of SRI:

> **Hypothesis 3**: Participants who see a consistently elevated SRI from their partner will rate their partner more negatively on mood attributes, compared

to participants who see a consistently normal SRI in uncertain and risky social interactions.

Hypothesis 4: Participants who see an elevated SRI will have lower (4a) trust rates (4b) cooperation rates in uncertain and risky social interactions compared to participants who see a normal SRI.

Experimental Design and Methods

The second study was identical to the heartrate study in every way, except that we told participants we were measuring "Skin Reflectivity Index," instead of heartrate. All mentions of the word "heartrate" in our original experiment software were replaced with "SRI" and/or "Skin Reflectivity Index". We purposely did not define or explain what the SRI signal is, or what its measurements mean. All participants were debriefed on the true nature of the experiment at the conclusion of the study. This debriefing included the fact that the partner was based on idealized behavior, and "SRI" was actually just a term for heartrate, as collected by a standard light-based pulse sensor. As with the first study, participants had the ability to ask the experimenter questions at the end of the study, or send an email if they had additional questions or concerns. We did not receive any follow-up concerns from participants. The only other variation from the first experiment is that, in the SRI experiment, we told participants to place their palms an inch above the light sensor rather than to place their fingers on the monitor. Since placing a finger on a light sensor is a familiar of measuring heartrate, this was done to reduce the possibility that participants would think that SRI is actually heartrate.

Participants

We recruited our sample for the second study from the same population and using the same method as described in study 1. Our recruitment procedures ensured that no one who participated in the first study could be recruited for the second study. Sixty-three participants (63) completed the second experiment, 40 women, 22 men, and one self-identified as 'other'. The mean age of participants was 21. Importantly, the gender distribution and age of the sample was equivalent to the first study.

4.5   Results

Quantitative results

H3 predicts that when individuals believe that their partner has a consistently elevated SRI, compared to a normal SRI, they will rate the partner more negatively on mood attributes. As with the first study on heartrate, we found an overall strong, statistically significant effect and medium practical association between attributions and experimental condition, $F(4, 59)$ = 4, p < .01; Wilk's lambda = .79, partial eta squared =.21. For the individual outcomes,
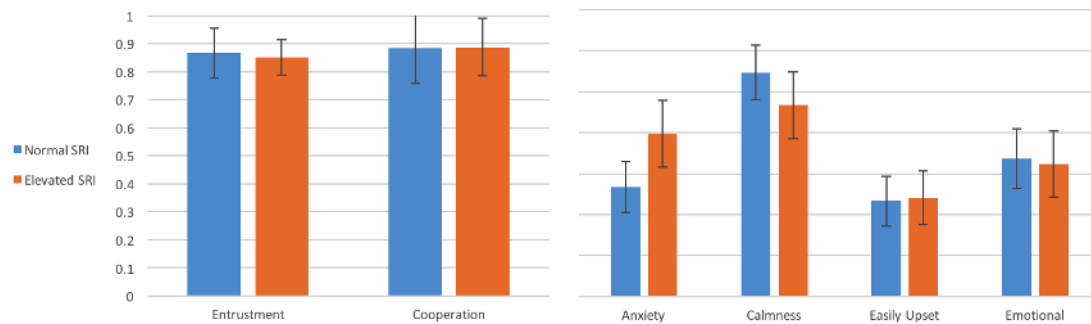
Figure 4.4: Means of entrustment and cooperation (left) and mood attributions (right) in elevated and normal SRI conditions.

we find that perceptions of the partners' anxiety is significantly higher in the elevated SRI condition (M = 3.97, SD = 1.62) compared to the normal SRI condition (M = 2.67, SD = 1.24), $F(1, 62) = 12.8$, $p < .001$; partial eta squared = .17. Furthermore, participants rated their partners as significantly more calm in the normal SRI condition (M = 5.5, SD = 1.3) compared to the elevated SRI condition (M = 4.68, SD = 1.63), $F(1, 62) = 4.4$ $p < .05$; partial eta squared =.07. Just as with the heartrate study, we found no statistically significant differences for perception that the partner is 'easily upset' or that the partner is 'emotional' (p = n.s.). In sum, we find strong statistical and practical differences in perceptions of both anxiety and calm, but no statistical or practical differences in how emotional or easily upset one perceives the partner to be in SRI conditions. Given the significant omnibus test and significant results on two of the 4 individual outcomes, Hypothesis 3 is partially supported.

Our final hypotheses predict that participants in the elevated SRI condition will exhibit lower trusting (H4a) and cooperative (H4b) behavior compared to those in the normal SRI condition. The average points entrusted by participants in the elevated SRI condition (M = 8.5, SD = 1.27) was not significantly different than the normal SRI condition (M = 8.7, SD = 1.77), t =.39, p = n.s, one-tailed test. Thus, individuals entrusted points to their partners at approximately the same level in both conditions (Figure 4.4). Unlike the heartrate study, however, we found no significant difference in cooperation rate between in the elevated SRI (M = .89, SD = .21) and the normal SRI condition (M = .88, SD = .25), t = .09, p = n.s., one-tailed test. H4a and H4b are not supported.

As with the first study, the simulated actors in study 2 were programmed to be consistently trusting and cooperative in the elevated and normal SRI conditions. Thus, we do not expect participants to rate the simulated actors differently in terms cooperativeness and trustworthiness between experimental conditions. As expected, the omnibus test of difference in perceptions of the trustworthiness and cooperative behavior between conditions was not significant, $F(2, 61) = 3$, p = n.s.; Wilk's lambda = .91, partial eta squared =.09.

Qualitative results

As in the heartrate condition, participants in the SRI condition were asked open-ended questions at the end of the post-experiment questionnaire, before the demographic questions and debrief. As in the heartrate condition, participants were asked how they would describe their partner. However, unlike in the heartrate condition, participants were asked, "Recall what we were measuring with the sensor. Please describe it below." After completing this question, participants proceeded were given two more open-ended items: "What, if anything, did SRI (skin reflectivity) tell you about your partner during this experiment?" and, "As a signal, what do you believe that SRI says about another person?"

The Meaning of an Unfamiliar Biosignal

We purposely did not explain what SRI might mean in this study. Nevertheless, when asked what was being measured in SRI, some participants gave us thorough explanations: The "reflectivity" part of SRI leads me to believe that the device is measuring how much light is reflected by a person's palms, which leads me to assume that SRI is increased when a person's hands are sweatier, and thus more covered in water, which reflects light better than simply someone's skin.

While explanations like this one indicate that participants believed our signal was real, reports of what participants thought SRI meant in the context of the game are more relevant to our analysis here. Like in the elevated heartrate conditions, and elevated SRIs were associated with either nervousness or excitement.

> If the SRI reads high, it may indicate that the person expects to be betrayed in some way or is hopeful of a positive result. I forgot what SRI stands for again. Since his/her SRI is always elevated, I would assume he/she is nervous/excited or just it's hot in here.

> SRI may give insight as to how nervous or excited someone's response is to something that happens. Maybe someone with a larger range in SRI is more emotional.

These assessments of SRI are quite similar to interpretations from the elevated heartrate, and corroborate our quantitative findings that those who saw elevated SRI rate their partners as more nervous. However, the fact that these emotional assessments were similar in both elevated heartrate and elevated SRI conditions, but behavioral outcomes were different, challenges our notion that negative emotional cues caused these behavioral outcomes—a point we address in more detail in the discussion below. As in the heartrate conditions, some participants responded that SRI told them little or nothing of interest about their partner:

> Nothing at all about the person other than an arbitrary value of a sensor.

> Since the SRI seemed to be bouncing around in the blue range but never got into the red range (which I assume would be "abnormal" since the blue range was normal) I don't think SRI is an accurate measurement of much.

As with heartrate, people cannot always be convinced that a biosignal is informative, even after many rounds of conditioning and a highly suggestive context. However, as in the heartrate condition, responses indicating that SRI had no meaning were a clear minority in our sample.

Elevated SRI

To help explain why elevated heartrate had a chilling effect on cooperative behavior, where elevated SRI did not, we delve into the responses of participants in the elevated SRI condition. When asked what SRI told them about their partner, participants often reported nervousness or anxiety, just as we noted in the quantitative results:

> [SRI shows] stress or heightened anxiety
>
> how reactive they are, or how close to the surface their emotions are.
>
> The nervousness of a person.

However, we noted that a significant number of participants in this condition mentioned that elevated SRI had some sort of positive association with behavior—even though it is also interpreted as indicating anxiety.

> Elevated means they feel safe and trustful. Lower than average means they are defensive and scared.

This interpretation stands in stark contrast to elevated heartrate, which also signaled anxiety, but had a negative association with behavior. In explaining why participants found elevated SRI to signal cooperativeness and trust, we look toward the responses of participants who seemed to learn a meaning for this signal:

> Well, since their SRI was always high and they always gave the money back to me, (based on these only two bits of info I know) I assume the two are correlated and an elevated SRI means that they're going to give the money back. [...] I guess it means that they're trustworthy and will do the right thing by their partner.
>
> I cannot tell [what SRI means], but my partner's was extremely elevated for the whole experiment and s/he was good at conducting mutually beneficial transactions.

These quotes strongly suggest that, unlike for heartrate, SRI participants picked up on a pattern between their partner's always-cooperative behavior and the elevated biosignal that we displayed to them, thus filling in the gaps about what SRI meant in this context. In contrast, we found no evidence that elevated heartrate participants learned such an association in the first study, despite the fact that every participant interacted with a perfectly cooperative partner in all conditions and studies.

Normal SRI

As with those in the elevated SRI condition, many participants in the normal SRI condition identified some relationship between SRI and the other person's mood. I think this helps identify how people are feeling internally when making decisions.

> his/her mood at that point of time
>
> [SRI shows] stress or heightened anxiety
>
> how anxious they are.
>
> I think our anxiety is being measured.
>
> How anxious/nervous someone is, if their SRI is high

In some cases, participants in the normal SRI condition inferred that elevated SRI might have a negative meaning:

> not to sure, high sri may indicate panic/fear or anger low sri may indicate calmness and contentness.
>
> A person is less likely to trust other people if he or she has a high SRI.

Overall, the responses for both SRI conditions support the interpretation that participants learned an association between cooperative, trustworthy behavior from the partner and SRI. As we argue in the following discussion, such associations are more likely in the SRI conditions because, unlike for heartrate, participants should have no preexisting beliefs or associations with SRI.

Limitations

Controlled, laboratory studies always come with clear advantages (such as high internal validity) and disadvantages (such as reduced external and ecological validity). Our study did not attempt to emulate a real-world interaction context with a biometric sharing device, though this is a clear next step, now that we know there are important differences in how biosignals are interpreted. Furthermore, our use of highly cooperative, computer-controlled interaction partners with stable biosignals (always high or always normal), prevents us from being able to speak to the effects of more dynamic behaviors and/or changes in biosignals

over longer periods of time. From these experiments, we also do not know how these results will transfer to other contexts, and other types of social interactions. Also, our study by nature focused on first-time, iterated interactions, both with an interface and with another unknown person. We do not know how these results might apply over the course of more personal relationships, or after repeated experiences with a specific interface in a biosignal sharing device. In addition, this research was conducted on young adults at a large public university, which is an important limitation when considering whether these results would hold across age groups and other key sources of sociodemographic variation in the larger population.

4.6   Discussion

We found that both heartrate and SRI signaled negative mood to participants, including anxiety and lack of calmness. It is possible that almost any "elevated" biosignals could be associated with negative mood attributions such as anxiety and lack of calmness: many elevated signals (pulse, temperature, blood pressure) carry associations with being angry, sick, hot-headed, and a host of other negative attributions. People may default to such attributions when seeing an unknown signal that comes from the body.

Elevated heartrate had a chilling effect on cooperation, where an unfamiliar biosignal, SRI, did not. So, why did the negative mood attributions in the elevated SRI condition not translate into reduced cooperation, as they did for elevated heartrate?

Our results shed light on two relevant phenomena that may address this question. First, pre-existing beliefs about heartrate are powerful: even when playing with a very cooperative, trusting game partner, negative connotations surrounding elevated heartrate appear to lead individuals to cooperate less. Our results suggest that participants bring to uncertain social interactions their own expectations about what elevated heartrate means, and that these biases cannot be quickly overridden, even when behavioral evidence sends a positive message (e.g., high cooperation and trust from the partner).

Second, we find evidence that participants can "learn" a social meaning for a previously unknown signal. Our qualitative data suggest that participants in the SRI condition associated whichever signal they saw (elevated or normal) with cooperativeness, and trustworthiness. Unlike with heartrate, people did not have preconceived notions of how SRI should affect the social behavior of the partner, since SRI does not exist. Instead, we observe participants discovering "what SRI means" by watching their partner's behavior in relation to the biosignal. In the absence of guidelines for interpreting what SRI is or what it measures, individuals appear to fill in the gaps with available behavioral information.

Our observation that people can learn social meanings for previous unknown signals begs a related question: Can pre-existing connotations for familiar biosignals change over time? The meanings of a signal like heartrate are the product of associations that have been shared and developed over centuries. However, technology allows for new expressions of these ancient signals [93]. If social heartrate information became an easily accessible

biosignal in trust-based interactions like negotiations, we might find its social meaning could evolve further. Unfortunately, short-term laboratory studies such as this one are unlikely to trigger or detect enduring shifts in the social meanings of familiar biosignals. We need both longer-term experiments, and mixed-methods research that can draw from rich qualitative data as well as statistically and practically significant changes in interpretations over time.

Broadly, our results raise questions about how and why unfamiliar signals take on social meanings in different contexts of interaction. Researchers in CSCW and HCI have long noted our tendency to read into cues and signals in computer-mediated communications. From impact factors and citation counts in scholarly work [36] to societal indices [102], to health metrics such as the bodymass index (BMI) [16], human have a tendency to impart "real" meanings onto metrics, scales and signals – meanings that may not align with the concepts their designers aimed to measure. It is critical that we continue to question how biosignal data could shape our interpersonal interactions, and whether the outcomes will always translate into meaningful social information.

4.7 Implications for design

From research projects like the sociometer, which produce "social metrics" [105], to consumer devices like the Spire, which compute "calmness" or "focus" quotients [96], developers are throwing different biometric signals at people faster than they can learn what the signals mean in context. In the absence of strong cultural beliefs about the signal, people could produce correlative assumptions similar to the ones we observed in our experiment. Designers should take care to establish what the signals in the applications mean, or could mean. Testing the limits of what people are willing, or able, to believe, and whether these beliefs transfer between different contexts, could have wide-reaching implications for those who design interactions with wearable biosensors.

On the other hand, many research and commercial projects use signals that people might associate with commonly understood experiences (e.g., a racing heart, a sweaty palm). Designers should strongly consider how these embodied experiences might color the conclusions that users jump to, and bound what users are willing to believe.

We also hope that researchers will investigate settings in which biosignals vary over longer time periods, perhaps with a more naturalistic technology probe study. Such a study could help us understand how prior beliefs about signals both affect and are affected by social interactions in the course of everyday life.

In general, wearable sensors can enable social interactions in which we share more information than is normally possible face-to-face. The ability to surface signals that are normally socially invisible (e.g. heartrate, or galvanic skin response) presents new territory for designers of computer-mediated interactions. Future work should continue to explore deeply how these novel signals fit into our existing understanding of social cues [50].

4.8 Conclusion

We find that sharing heartrate can negatively influence trusting attitudes and behaviors. 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.

Further, our experimental results imply that interfaces can "teach" the meaning of some biosignals, where others carry strong, pre-existing connotations that even repeated interactions cannot easily alter. In general, prior beliefs about the body (drawn from culture, lived experience) seem to shape what a biosignal can mean in a given context. However, in the absence of prior beliefs, there exists an opportunity—and a potential danger—that designers of biosignal-sharing systems can condition participants to learn (potentially arbitrary) associations between biosignals and social behaviors.

Aside from heartrate, we do not know which of many other biosignals might be associated with moods and behaviors. Other biosignals (e.g., galvanic skin response, electroencephelography 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. Similar studies with signals from, e.g., the brain [4] are a clear direction for future work. Especially interesting cases are signals for which precise or empirical meanings are still being hotly debated, such as EEG (brainwaves), a sensing modality we begin to discuss in the next chapter.

Chapter 5

Shifting to the brain

While the prior chapter establishes that people build mind-related meanings around biosensory data, this chapter locates brainscanning as a fruitful case for understanding how particular sensing technologies construct notions of mind. I report on the qualitative and quantitative results of survey among participants in a large (n>10,000), longitudinal health study, and an Amazon Mechanical Turk population.

What can different biosensors reveal about what you are thinking and feeling? In this study, we posed this question to 200 people, half of whom came from Mechanical Turk, and half from a longitudinal study in which subjects contribute sensor data to track health outcomes. We were interested in how people perceived risks around the disclosure of sensor data, and how their expectations related to both the type of device in question, and the participants' prior experience with disclosing data from wearable devices.

Through a quantitative and qualitative analysis of survey data, we find some differences in perceptions of risk between populations. However, we find that certain devices draw greater notions of risk of mind-reading than others. In particular, electroencephalography (EEG) appears to carry an unusually high perceived risk, beyond even fMRI, which has proven more revealing in past studies [58]. We discuss implications for the design of EEG-based brain-computer interface, a modality rapidly gaining in popularity in the technology industry [64, 75, 72], and for wearable technologies generally.

5.1 Background

In their qualitative study of activity trackers, Rader and Slaker (2017) found that the "visibility" of tracking devices (how data are measured, and what data are calculated as a result) has a large impact on the way people understand these devices as working, and may impact the privacy decisions users make as a result [85]. While this study looked at a broad array of sensors, it did not study particular threats to privacy. Meanwhile, past work in CSCW and beyond has demonstrated that people build meanings around shared data from wearable sensors pertaining to mood, emotions, and other aspects of mind [69].

These studies raise the notion of the mind as a site for exploring perceptions of sensor data, and what these data might mean. However, the interpretations surfaced by previous studies are typically contextual, specific to particular social contexts [98], and to particular types of sensors. However, it is not clear from these studies how different sensors compare to one another in the way users assess the risks of data disclosure.

In this work, we aim to study a specific privacy threat (knowing what a person is thinking and feeling) across a variety of sensors. Through quantitative and qualitative data, we aim to perform inductive work around two preliminary questions: (1) Which sensing devices seem the most (and least) likely to reveal what a person is thinking and feeling? (2) How do these perceptions change according to this person's observed willingness to share sensor data with others? In the following section, we outline how we examined these questions using a survey, deployed across two distinct populations.

5.2   Methods

Our survey consisted of a question in which subjects ranked various sensors: "Please rank the following sensors in how likely you believe they are to reveal what a person is thinking and feeling." Our selection of sensors (Table 5.1) aimed to include both sensors commonly found in wearable and mobile devices, and sensors more commonly associated with the medical industry. We sought to achieve a mix of modalities found only in medical devices, found only in commercial devices, and found in both commercial and medical devices.

Data	Medical?	Commercial?
Facial expression	No	Yes (camera)
Body language	No	Yes (camera)
Brainwaves (EEG)	Yes	Yes
Eye movement	No	Yes
Heartrate/pulse	Yes	Yes
MRI/fMRI	Yes	No
Blood pressure	Yes	No
Skin conductance	Yes	Yes
Blood oxygenation	Yes	No
Step count	No	Yes
GPS + accelorometer	No	Yes
VR headset	No	Yes

Table 5.1: Sensors referenced in the survey.

To capture a population willing to share sensor data, we submitted our survey to participants in Health-e-Heart, a large (n > 40,000) longitudinal study in which subjects volunteer to share data from wearable sensors longitudinally so that researchers may monitor health outcomes [37]. To compare this population to a more general population, we also submitted
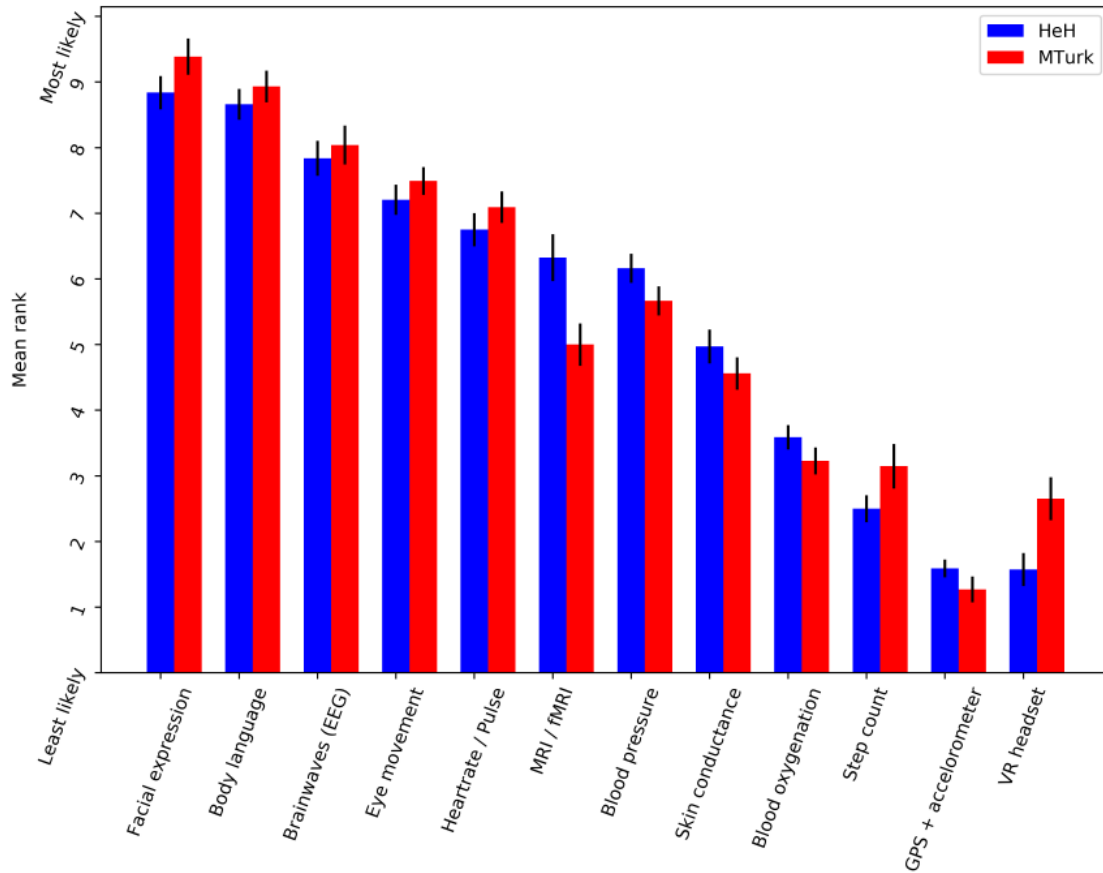
Figure 5.1: "Please rank the following sensors in how likely you believe they are to reveal what a person is thinking and feeling." Higher bars indicate higher rank, or higher likelihood of being revealing.

our survey to Mechanical Turk workers in the United States. Our survey included 100 Health-e-Heart participants and 100 participants from Mechanical Turk.

5.3   Results

Quantitative results

In our rankings, brainwaves (EEG) are seen as among the most revealing biosignals, just below body language and facial expression, in their capacity to reveal the inner workings of a person's mind. More common sensors such as GPS and step count are seen as less revealing (despite empirical evidence suggesting such data can be quite revealing indeed [17]). Mechanical Turk participants thought virtual reality headsets and step counters were significantly more likely to reveal what a person is thinking and feeling than did Health-e-Heart subjects. On the

other side, Health-e-Heart subjects believed fMRI, blood pressure, blood oxygenation, and GPS/accelorometer were significantly more revealing than did Mechanical Turk participants.

Qualitative results

When we asked subjects to reflect on why they answered the way they did during the ranking task (Figure 5.1), EEG solicited the strongest and most diverse reactions. Since this sensing modality is still relatively obscure in consumer devices, we delved more deeply into qualitative data in hopes of explaining these concerns. Subjects in both groups generally believed EEG to reveal various details about the mind, mood, emotions, and identity. In the Health-e-Heart group, several subjects gave relatively specific explanations as to why they ranked this sensing modality highly.

> *(S24) I assume some information can be gleaned from brain wave activity in various parts of the brain related to rewards or executive control, but without accompanying information, it may be difficult to discover my thoughts.*

> *(S23) EEGs note parts of the brain that are active. Again, in conjunction with other measurements, I suspect that some sense of what one is thinking and feeling could be learned.*

> *(S91) I would rate this relatively high on the list because science has shown that we can detect a lot about which areas of the brain are accessed and at which times. This can tell a person a lot about what they might be thinking and especially how they are feeling.*

While these explanations range somewhat in their specificity and confidence, they share the general sentiment that EEGs can be revealing. Subjects in the Mechanical Turk condition broadly shared this belief, though tended to use less physiological detail in their explanations.

> *(S157) Brain activity can pinpoint exact emotions by monitoring certain areas on the brain.*

> *(S130) Brainwaves could tell you a lot more about what someone is thinking and feeling. You could measure the patterns of brainwaves in an experiment.*

Meanwhile, some subjects from both groups did not fit this trend. Ten subjects ranked EEG low in its ability to measure what a person is thinking or feeling. Their qualitative answers revealed a diverse set of reasons for this ranking. Three subjects indicated a general lack of faith in brainwaves' reliability.

> *(S20) I don't think we have the ability to translate brainwaves into thoughts or emotions.*

> *(S101) EEG is very nonspecific and rarely can tell details reliably.*

> *(S138) Possible but not accurate.*

These explanations broadly centered around EEG as a signal. They range somewhat in their confidence, from a fundamental skepticism (S20) to caveats about possible accuracy or specificity (S101, S138). In contrast to these three subjects, S10 ranked EEG low because s/he felt the premise of a consumer grade EEG was implausible.

> *(S10) I assume that scientists can identify by brain patterns what others are feeling and thinking based off of years of research. I've never heard of a consumer grade eeg - and doubt it could be as powerful as a laboratory eeg. If it is then I would be interested in this product.*

This subject's explanation surfaces the practical differences in attitudes that people might have to a technology's theoretical existence, and its realized existence as a consumer device. Future work could look more closely at how the presumed scientific authority of a brainscanning apparatus affects people's willingness to accept specific BCI applications. Finally, one subject's skepticism what brainwaves can reveal stemmed from his/her personal medical experiences.

> *(S116) My son has absence seizures, so his brainwaves change.*

This particular quote highlights how individuals' life experiences might shape the way they engage (or refuse to engage) with brain-sensing devices. In general, this quote and others motivate the need for a rich, qualitative understanding of people's first-hand experiences with brainscanning devices, as well as data collection, in order to understand what role BCI applications such as passthoughts could play in day-to-day life.

5.4 Discussion

Our results find some differences between the Health-e-Heart and Mechanical Turk groups, particularly around devices with medical associations. However, device rankings were mostly the same between conditions. Our findings indicate that sensing modalities play a large role in building understandings of what sensors might reveal, along with prior experiences sharing sensor data. We discuss implications for design in sensor-based interactions: different sensors may trigger different concerns about privacy, which could in turn trigger debates about what counts as a valid privacy concern, and what does not.

Health-e-Heart participants believed fMRI, blood pressure, and blood oxygenation to be more revealing than participants in the Mechanical Turk condition. Since these subjects are participating in a medical study, it is possible that they are more attuned to what medical devices can reveal, or simply that they are primed to think about them. Health-e-Heart subjects also thought that GPS and accelerometer were more revealing than their Mechanical Turk counterparts. This differences indicates that the HeH subjects' constant participation in monitoring does not make them less sensitive to privacy concerns (i.e., they do not "acquiesce"

to such monitoring). It does perhaps suggest that their knowledge of tracking modalities differs, a suggestion supported by our qualitative analysis.

Conversely, Mechanical Turk participants believed the VR headset and step count were more revealing than did the Health-e-Heart subjects. We found no significant difference in experience with virtual reality between the two groups. Future work should examine possible causes for this difference. As virtual reality grows in popularity, and as the producers of these devices increasingly attempt to outfit VR headsets with sensors [65], it will be important to understand what about VR causes people concern.

It is worth noting that Mechnical Turk participants may be subject to monitoring as well, as the human-intelligence tasks they perform on the platform may subject them to various types of surveillance (e.g., clicks, timing activity, question checks, browser fingerprinting, etc). Future work should examine more deeply Turker's knowledge of, and response to this sort of tracking, issues which connect to to broader questions of digital surveillance in the workplace.

Our most surprising finding, consistent across both groups, was the overall high ranking of EEG. EEG was perceived as more likely to reveal what a person is thinking or feeling than fMRI, which prior work indicates to be a more detailed brainscanning apparatus [58]; EEG is course-grained in comparison. Future work should examine more closely why EEG was so highly ranked (e.g., perhaps participants did not know what fMRI is). Reasons aside, EEG's high rank in our finding offers both opportunities and challenges for designers. People's belief in EEG's ability to sense intimate details may allow designers to create creative, helpful or therapeutic applications [54]. On the other hand, these same beliefs could allow designers to trick users [4], or might dissuade prospective users from wearing EEG at all. These questions are increasingly important as EEG-based BCI is gaining interest in industry [72, 75] and in the public imagination [64, 99]. How will people encounter these devices, and find their data meaningful (or not) in the course of life? The answer to these questions depends heavily on what users think their data can reveal. Thus, future work should look longitudinally at EEG and BCIs as these devices ebb and flow in the public (and corporate) imaginary.

Implications for design

Our studies reinforce past work in demonstrating the relevance of everyday theories in understanding what sensors can reveal [85]. However, our work also indicates that different sensing modalities may heighten particular privacy concerns (e.g. EEG). By the same token, other devices may obfuscate privacy concerns, creating a compromising position for users as they are lulled into a false sense of security. For example, GPS and accelorometer have together been used to detect mental health status [17]; the fact that these sensors were not rated highly gestures toward differing concerns across sensing modalities, and the fact that these concerns may not align with technical efforts among designers and engineers. In general, future work should examine more deeply how prior experience with devices meets with expectations about the body to produce understandings of privacy, what devices can "know" (and what counts as knowing). As emerging devices (such as VR and EEG) become more familiar to users, future work should monitor beliefs about sensing modalities as these

technologies develop. Sensors such as GPS and accelorometer are now ubiquitous, but attitudes around them have likely changed since their introduction [27]. Through longitudinal studies, we stand a chance at observing changes in attitudes, thus putting us in a position to anticipate changes in privacy attitudes and privacy-preserving behaviors.

5.5   Conclusion

Our findings complicate recent work around the folk interpretations of sensor data, indicating that prior experience with sensors is only one way to understand where interpretations of sensor data come from. Beliefs about the body play an important role in shaping beliefs about what sensors can know. As industry pushes toward new sensing modalities such as EEG, future work should remain critical in probing the beliefs of end-users, as their apprehensions will shape the sorts of applications that users are willing to accept.

Chapter 6

Talking to engineers about brain-computer interface

As we saw in the previous chapter, EEG triggers intriguing beliefs about the knowability of the mind. In this chapter, we use EEG to shift from users of sensing devices to their engineers. Having motivated EEG as a case study for further exploration, this chapter examines the beliefs of software engineers through their interactions with a working brain-based authentication system. This population's beliefs are particularly critical as consumer brainscanning devices have become open to tinkering through software. Although we find a diverse set of beliefs among our participants, we discover a shared understanding of the mind as a physical entity that can and will be "read" by machines. These findings shed light on what sorts of applications engineers may accept as buildable, and prime our concluding chapter on how built artifacts may come to structure our notions of what minds are.

6.1   Background

In 2017, both Mark Zuckerberg and Elon Musk announced efforts to build a brain-computer interface (BCI) [64]. One blog post enthusiastically describes Musk's planned BCI as a "wizard hat," which will transform human society by creating a "worldwide supercortex," enabling direct, brain-to-brain communication [99].

A slew of inexpensive brainscanning devices underwrite such utopian visions. 2017 saw a BCI for virtual reality gaming [75] and brainwave-sensing sunglasses [94] join the already large list of inexpensive, consumer BCIs on the market [64, 54, 44]. These devices, which are typically bundled with software development kits (SDKs), shift the task of building BCIs from the realm of research into the realm of software development. But what will software developers *do* with these devices?

This study employs a technology probe to surface narratives, and anxieties, around consumer BCIs among professional software engineers. We provided a working brain-computer interface to eight software engineers from the San Francisco Bay Area. As brainscanning

Figure 6.1: A participant uses our brainwave authenticator in his startup's office.

devices become more accessible to software developers, we look to these BCI "outsiders" as a group likely to participate in the future of brain-computer interface. Specifically, we provided participants with a brain-based authenticator, an application predicated on the notion that a BCI can detect individual aspects of a person, making it a potentially fruitful window into broader beliefs about what BCIs can reveal [87, 35].

Despite heterogeneous beliefs about the exact nature of the mind, the engineers in our study shared a belief that the mind is physical, and therefore amenable to sensing. In fact, our participants all believed that the mind could and would be "read" or "decoded" by computers. We contribute to an understanding of how engineers' beliefs might foretell the future of brain-controlled interfaces. If systems are to be built that read the mind in any sense, we discuss how such systems may bear on the long-term future of privacy and cybersecurity.

Brain-computer interfaces & pathways to broader adoption

BCIs allow people to interact with computers without muscular action. Instead, nervous system activity is translated to a discretized (digital) signal. BCIs can be categorized broadly as invasive (requiring implantation) or non-invasive (requiring only external, removable

equipment). Non-invasive, consumer BCIs, are lightweight, require minimal setup, and do not require special gels. EEG (electroencephalography) is currently the most viable choice of sensing modality for consumer BCIs [19].

Historically, researchers have conceived of BCIs as accessibility devices, particularly for individuals with severe muscular disabilities. However, accessibility devices can sometimes provide routes for early adoption, and thus broader use. Speech recognition, for example, was once a tool for individuals who could not type; eventually, it became adopted as a tool for computer input, now commonplace in IoT devices such as Alexa and Siri. Since accessibility devices can give rise to broader consumer adoption, we ask what such a pathway might look like for brain-computer interfaces. With an expanding array of inexpensive brainscanning hardware, many of which come bundled with engineer-friendly SDKs, the pathway to a future of consumer BCI increasingly becomes a matter of software engineering.

Thus, we look to software engineers in the San Francisco Bay Area. We use these engineers as a window into broader beliefs about "Silicon Valley," a term we use here to stand in for the technical, economic and political climate that surrounds the contemporary technology industry in the area [89]. While we do not believe only Silicon Valley engineers will influence the future of BCIs, historically, these engineers have an outsized impact on the types of technologies developed for mass consumption, especially with respect to software. As BCI hardware becomes more accessible, and therefore more amenable to experimentation as software, this group once again holds a unique role in devising a consumer future for this biosensor. Indeed, the Muse, and similar devices, have robust SDKs and active developer communities that are building and showcasing BCI applications [76].

However, we did not want our subjects to have first-hand experience in developing BCIs, as we did not want them to be primed by existing devices' limitations. Instead, we selected individuals who indicated they would be interested in experimenting with consumer BCI devices in their free time. This screening was meant to draw subjects likely to buy consumer devices and develop software for them. We believed that these engineers' professional expertise in software development afford a desirable criticality around our technical artifact.

What brain scans can tell

Brain scanning holds a unique *charisma* [5], not only among researchers in related fields [87], but among non-experts as well [4]. Ali et al (2014) found university undergraduates believed a brain scanning device (a fake one, unbeknownst to them) could reveal intimate details of their thoughts, even after receiving a lecture about the limitations of brain scanning technologies [4]. In that study, participants saw scans of the *brain* as informative with regard to the *mind*, a distinct entity that is potentially more expansive than the brain [25, 48].

This entanglement of mind and brain has been explored by past work in science and technology studies. For example, Dumit's (2004) study of positron emission tomography (PET) explores utopian (and dystopian) visions of diagnosing mental illness, or even criminality, from scans of a person's brain [35]. The idea of the mind's "legibility" via computational technologies has been concretely explored by Rose (2016) [87], who ties together a number of efforts across

neuroscience and cognitive science to argue that specific technical implementations from these fields (along with their rhetoric around, and beliefs about the brain) allow *the mind* to be "read" or "decoded."

However, there exists an opportunity to investigate how pervasive such beliefs are among those who are not neuroscience experts, yet nonetheless technical practitioners. Given the recent shift of brain scanning equipment from research tool to consumer electronic device, we ask what software engineers, newly able to develop applications around brain scanning, might build. Answers to this question could have far-reaching consequences, from marketing, to entertainment, to surveillance. In particular, we aim to center how engineers' ideas about the mind, especially its relationship to the brain and body, inform and constrain their beliefs about what BCIs can (and should) do.

A BCI technology probe

In this study, we use a technology probe to examine the beliefs of software engineers about what BCIs can reveal about the mind. Technology probes are functional apparati intended to both collect data *in situ* from participants, and to inspire participants to reflect on the probes, and on their beliefs more generally [53].

Probes have a long and diverse history within HCI, often referring to a variety of different practices [11]. In the context of our study, our probe seeks primarily to answer research questions, rather than to figure as one step in an iterative design process. Unlike some probes in past work ours was not intended for longitudinal deployment. Instead, we aimed to gather beliefs about particular technologies and domains through a session of open-ended interaction with a device [63].

Our probe's unfinished appearance was intended to invite critique and playful experimentation [32, 63]. However, unlike a mock-up or provocation, our probe did function as advertised, allowing participants to interact with the devices in an exploratory and unconstrained way (indeed, many engineers tested to confirm that the device's feedback was real). We designed our probe to steer participants away from providing narrow feedback about the interface at hand, and toward sharing their broader beliefs about the brain and mind.

Brain-based authentication

Our study employs a brain-based authenticator as a research probe to elicit engineers' beliefs about BCIs (and the mind and/or brain they purport to sense). This section explains how brain-based authentication works, and why we chose this application for our study.

Authentication (i.e., logging into devices and services) entails a binary classification problem: given some token, the authenticator must decide whether or not the person is who they claim to be. These tokens typically relate to one or more "factors": knowledge (something one knows, e.g. a password), inherence (something one is, such as a fingerprint), or possession (something one has, such as a device) [24]. Brain-based authentication relies on signals generated from individual's brains to uniquely authenticate them, which has a

number of potential advantages over other authentication strategies (see [71] for a review). First, brainwaves are more difficult to steal than biometrics fingerprints, which are externally visible, and left in public as one's hands touch objects in the environment. Brainwaves also change over time, making theft even less likely. Second, brain-based authentication requires no external performance, making it impervious to "shoulder-surfing attacks" (e.g., watching someone enter their PIN).

We chose to build a brain-based authenticator for our study for a few reasons. First, having participants use a functioning system helped them imagine how they might use BCIs themselves. Second, the system is a plausible one, backed by peer reviewed research, thus we expected our participants to judge its claims credible. Third, the system embeds particular assumptions about what brain scanners are able to capture. Our system embeds ideas that our Muse headset can capture aspects of individual brains that are unique; as such, we expect that a working, brain-based authenticator will encourage participants to reflect not only on how a BCI applications might be adopted by the broader public, but also on what BCIs may be able to reveal about the mind and brain, and to critically examine the limits of what BCIs in general are able to do.

6.2   Building the BCI authenticator probe

Implementation

Since we wanted our technology probe to appear portable enough for use in the real world, we decided to use a pre-existing consumer EEG device to build our authenticator. We settled on the Interaxon Muse (Figure 6.1), a $299 headband that can be worn easily, transmits data wirelessly, and requires no gel to maintain contact between the scalp and electrodes [54]. Using a system that required conductive gel would have signaled to the participants that the technology is still limited to lab settings, and not yet ready for the real world, which could have influenced their responses.

Although the Muse's signal likely contains noise, a perfectly clean signal was not necessary to elicit beliefs from subjects in the context of our technology probe. Further, despite the Muse's small form-factor and dry electrodes, past studies have verified its signal is sufficient quality for some neuroscientific research [60].

Due to the device's battery life and intermittent connectivity when walking, the Muse headband did made a longer-term study impractical. Thus, we opted to perform a study over a short time and in a controlled environment, drawing on past technology probe studies with similar constraints [32, 55].

Data from the Muse was collected via the device's native OSC interface, and stored in a timeseries database. Queries from this database were used to provide training data for a machine learning classifier. In a preprocessing step, we performed a fast Fourier transform (FFT) to generate frequency-domain data from the time-domain data. In the machine learning step, we split a corpus of readings (and labels) into train and validation groups.

Figure 6.2: Our probe's visualization of 1's and 0's gave our engineers a "raw" view of the authenticator's behavior. Pictured, the UI (a) accepting someone, (b) rejecting someone, or (c) presenting mixed, ambiguous feedback.

Using XGBoost [21], we trained a binary classifier on seven different splits of the train group. After the classifier was produced, we validated its performance on the withheld validation set.

Given a target participant to classify, our classifier used any reading from this participant as a positive example, and any reading *not* from this participant as a negative example. Negative examples also included signals with poor quality, and signals from which the device was off-head or disconnected. Ideally, the resulting classifier should produce "authenticate" labels when the device is on the correct person's head, and "do not authenticate" labels at any other time. This classifier could output its labels to a simple user interface (UI), described in the next section.

Interface

As the device produces data, the classifier outputs labels of "accept" or "reject." Our interface displays these labels as a square of 0s and 1s, which filled up as data from the device rolled in (Figure 6.2).

Several considerations motivated this design. First, the UI represents the probabilistic nature of the classification process. Individual signals may be misclassified, but over blocks of time, the classifier should be mostly correct (represented as blocks of mostly 0s by our interface). Thus our simple UI makes visible both the underlying mechanism of binary classification, and its probabilistic nature. Second, because our UI provides potentially ambiguous feedback (as opposed to unambiguous signals of "accept" or "reject"), it allows for potentially richer meaning-making and explanatory work [91]. Toward this end, the UI's real-time reactivity ("blocks" of 1s and 0s filled in over time) allows participants to experiment actively with the device, forming and testing hypotheses as to what makes classification succeed or fail.

Finally, our UI gives the probe an "unfinished" appearance. We believed this interface would cause our participants to activate their "professional vision" as tech-workers [43], and critique or test the device as if it were a design of their own. Ideally, we hoped participants would intentionally stress-test the device, or find playful ways of misusing it. These misuses could allow participants to form hypotheses about why and how the device succeeds and fails.

6.3   Methods

We recruited participants by word of mouth. A recruitment email explained that subjects would interact with a working BCI, and be asked their opinions about the device, and about BCI broadly. We screened respondents by their current occupation and stated interest in experimenting with BCIs in their free time. All participants were employed full-time as software engineers at technology companies in the area.

A total of eight people participated, three of which were women. Participants' ages ranged from 23 to 36. We met with subjects for a single, one-hour session in which we trained and tested a brain-based authenticator, allowing them to interact with it in an open-ended way.

These sessions were designed as a semi-structured interview, interspersed with conversation between the researcher and the participant. Our study protocol was approved by our institutional IRB. Interviews were recorded, and later transcribed. We performed an "issue-focused" analysis of the transcriptions [101], allowing topics and themes to emerge during analysis. To protect subjects' anonymity, all names have been changed to pseudonyms.

Wearing the device

The interviewer began by explaining that participants would wear a BCI, which we would train to work as an authenticator, answering participants' questions about how the device works. Subjects were told that they would be asked about their opinions on BCIs generally, and that their anonymized voice and EEG data would be collected.

The interviewer asked participants to place the EEG headband themselves, and to assure that the device fits comfortably, at which point the interviewer would begin recording signals from the device. Next, the interviewer would ask participants how they felt about having the EEG device on their head. This question would typically begin a short, open-ended exchange about their past experience with brain-scanning devices, and prior knowledge, if any, of BCIs. This exchange would segue into a broader discussion about the participant's use and relationship with technology, in personal and work life.

After this initial conversation, the interviewer would perform a brief *calibration* step with the participant, in which data are collected to train a custom classifier for use in authentication. Participants would perform a number of tasks, or *mental gestures*, prompted by a stimulus presentation program. These tasks provide a more diverse corpus of an individual's signals, which should enable a more robust (and accurate) classifier. After this calibration procedure, which usually lasted about ten minutes, the interviewer would perform a semi-structured interview with participants. The interviewer would continue to record data from the Muse throughout this interview.

Using the authenticator

At this point, the interviewer would explain to participants that the data collected thus far would be used to train a custom authenticator for them. The interviewer would explain roughly how the authenticator would work: the probe should *accept* readings when the participant is wearing the device, and *reject* readings in any other case.

Next, the interviewer would run a script that trained our XGBoost classifier (Section 6.2). Participants could watch the training process run, if interested (a few were). After the training process completed, the researcher would set up the UI (Section 6.2) and allow participants to view the classifier's output in real-time using live data from the participant's Muse device. Participants would then see the probe's *accept* or *reject* classifications using the live data from their headset.

After allowing participants to acclimate to the output, and answering any preliminary questions, the interviewer would encourage the participant to experiment with the authen-

ticator, and share any impressions, reactions or ideas. The open-endedness of this session was meant to encourage participants to explore the device's capabilities and limitations, free of particular tasks to accomplish. However, we suspected that our participant population would be particularly prone to "hypothesis-testing," exploring the devices limitations by building theories about how it might work. We structured the session around this assumption, preparing to ask participants to think aloud as they explored the device's capabilities.

After some free-form exploration (usually involving some back-and-forth with the participant), the interviewer would transition into a semi-structured interview, which would occur with the device still active. The interviewer would ask participants to unpack their experience, and lead them to explore what they felt the device could reveal about them. After some discussion, the formal interview would conclude, and the participants would remove the Muse device from their head.

6.4 Experiencing the authenticator

In general, we found particular reflections to come at different points in the interview protocol. Critiques (and questions) about the device tended to come as soon as engineers placed the device on their heads. Reflections on the BCI broadly, and its future trajectories, tended to come after viewing the probe's feedback for some time. As these conversations progressed, participants naturally tended to reflect on what future BCIs might be able to do. Subjects would typically relate the capacities of the probe, and of possible future technologies, to their ideas about the mind, body or brain. The probe continued to run during these discussions. Toward the end of the interview, the researcher would prompt participants to reflect on any anxieties they might have about the future of BCIs (interestingly, only one participant raised this subject on their own). The remainder of this section is organized to roughly mirror this common order of participants' reflections during interviews.

Using the BCI probe

Our working authenticator elicited diverse reactions from the engineers in our study. Almost all participants cracked jokes after putting on the headband (three subjects commented that they felt like they were "from Star Trek"). All participants except Joanna said they would not wear the device in public, though a few conceded that they might if the headsets were more common. Terrance commented, "If most people are doing it, then it's fine. Sort of like stock speculation."

Perceptions of the authenticator's accuracy were mixed. Four participants found that the authenticator worked well for them. For these participants, the authenticator consistently rejected blocks when the headset was off of their head, or worn by the researcher (these participants had the idea to test the authenticator by asking the researcher to wear it).

On the other hand, four participants found the probe consistently rejected every reading, whether it came from them or the researcher (i.e., they experienced false rejections, but not

false acceptances). These subjects often tried to remedy the situation by attempting tasks they had rehearsed, typically with mixed success. Most of these subjects concluded that there was not enough training data to produce reliable classification, but that such a system would work with a larger corpus. In contrast, Alex, a 30 year-old founder of an indoor agriculture startup, blamed himself, saying "I must not produce very distinguishable thoughts."

Those participants who felt the probe's authentication was reliable tended to center their explanations on why it worked. Participants who experienced less consistent accuracy with the authenticator tended to center their explanations on how the device might be improved, e.g. with better or more comprehensive sources of data. This impulse to "fix" likely speaks to our participants' general tendency to engineer working systems.

As we hoped, the engineers engaged critically with the technical implementation of the probe. In general, engineers asked about the machine learning infrastructure underlying the authenticator, and several participants (particularly John, Mary and Alex) asked specific questions, and made specific recommendations, diagnosing issues with the authenticator by thinking about the diversity and size of the training set. Almost all participants noted the authenticator worked better when they were not looking at the visual feedback from the user interface. Participants generally theorized that this might occur because they were not viewing feedback when training the classifier. In these cases, the engineers appeared to apply their domain knowledge to their observations in using our technology probe.

Reflecting on the future of BCIs

Our technology probe caused almost all of our participants to speculate on the future of BCIs generally. To most participants, the future of BCIs seemed to be largely pre-determined. One of our participants, Terrance (a 24 year-old software engineer at a small transportation startup), removed the headband to inspect it, and commented on its awkward visibility. In doing so, he reflected on the future of BCIs, speaking in no uncertain terms about a future of computer-mediated "telepathy."

> Things just get progressively smaller until they disappear. And one day this'll just be an implant in my brain, doing crazy things. It'll be interesting socially, how people come to terms with it, when it's just an implant, or at least very pervasive ... I could send you a message, and it could be like you're thinking it yourself, even if you're on the other side of the Bay. *(Terrance)*

Terrance believed that BCI *will* become more prevalent: not just that smaller sensors will lead to more effective or usable BCIs, but that they will also result in greater uptake of the technology. While he references the social dimension of their adoption, he indicates that people will need to "come to terms with" the developments, rather than providing direct agency to users who may choose to adopt the technology or not.

Two participants felt less sure that such a future of pervasive BCI would ever come to pass. Elizabeth, a 30 year-old front-end engineer, noted skepticism about signal quality, or

usefulness outside of persons with disabilities. Mary, a 27 year-old software engineer at a large company, pointed to social reasons for her skepticism. In reflecting on the relative accuracy of the probe's authentication performance during her session, she commented that "90 plus percent" of people would be "totally freaked out" by brain-computer interfaces generally. She continued to say that companies may themselves stop BCIs from becoming too pervasive or advanced.

> I feel like those companies, even if this were feasible, there's a moral quandary they philosophically have not figured out. They will not let the research get that advanced ... I just don't imagine them being like, "okay computer, now read our brains." *(Mary)*

While the probe was effective in spurring subjects to talk about issues around BCIs, its accuracy as an authentication device did not seem to alter participants' belief in BCI's future as a widespread technology. Unsurprisingly, the four subjects who experienced reliable authenticator accuracy all expressed that BCIs would become commonplace in the future. However, only Joanna connected the device's poor performance in her session with a probability of ongoing accuracy issues for BCIs in the future. The other three subjects who felt the device did not perform accurately all offered explanations as to why, and explained that future devices would fix these issues.

Mind, brain, body

During their interactions with the probe, almost all of our subjects discussed their deeper beliefs about the nature of the mind, and its relationship to the brain and body. Since participants discussed the future trajectory of BCIs led to discussions while the probe continued to work (or fail), the subject often arose of what BCIs might be able to detect, even theoretically. As one example, John, a 26 year-old software engineer at a small chat startup, noticed that the authenticator only worked when he was speaking, but not when he was listening to the researcher. He offered an explanation for the discrepancy.

> There's probably some kind of fundamental difference between creating thoughts and consuming thoughts. You're still making thoughts, right, but it's almost like programming versus being programmed. *(John)*

When pressed on how strictly he meant his metaphor of programming, John confirmed that he meant it quite literally, saying, "I think we are just computers that are way more sophisticated than anything we understand right now." We return to this strictly computational account of the mind as "just" a computer in the discussion.

Mary gave a computational account of mind that was more metaphorical than John's, drawing on comparisons between machine learning and the mind. She cited the many "hidden layers" in deep neural networks, and that, like in the brain, "information is largely distributed."

While she believed deep learning models and the brain were "different systems foundationally," she said "there are patterns" that relate the two to one another, and indicated that advances in deep learning would spur a greater understanding of the brain.

Although six of our participants provided a largely computational account of mind-as-brain, not all did. Joanna, a 31 year-old engineer who previously completed a PhD in neuroscience, felt that the mind was "the part of the brain I am aware of, the part that is conscious." She believed that neurotransmitters throughout the body have a causal relationship to what happens in the mind, but do not constitute the mind themselves; the contents of mind occur physically in the brain, and the brain alone. In other words, her account is one of "mind as conscious awareness," and while unconscious phenomena affect mind (e.g. the body, environment), they are not part of the mind *per se.* Interestingly, the probe did not work well for Joanna, and she felt confident that its poor performance was due to contaminating signal from her body (a theory she tested, and validated, by moving around and observing the probe's feedback).

Meanwhile, in one subject's account, the mind extended beyond the confines of the body. Terrance felt that there was "no meaningful difference" between the body and brain, nor between the body and the physical environment at large, saying that "you can't have one without the other." He believed that all three of these entities constitute the mind in a mutually-dependent way. However, Terrance indicated that the mind is still strictly physical, as are these three entities. Although Terrance did not provide details on how exactly the mind extended beyond the body, it is interesting to note this position's similarities to Clark's (2013) account of the extended mind [25], or Hutchins's (2005) work on distributed cognition [52], though Terrance was familiar with neither.

Participants also offered differing levels of confidence in their beliefs about the nature of the mind. Joanna (who has a background in neuroscience) reported that "we do not know everything we need to know" about how the mind works. Three other subjects reported similar beliefs. However, those subjects with a computational account of mind tended to feel more confident that their account was substantially accurate.

> I think the consensus is that the body is mostly like the I/O of the brain. *(John)*

John's account here implies that a sufficiently high-resolution brain sensor would accurately capture all of a person's experiences. John confirmed this explicitly, saying, "if you could 3D print a brain, and apply the correct electrical impulses, you could create a person in a jar." In this computational metaphor of I/O (input/output), the body itself does not have agency; instead, the body actuates the brain's commands (output), and senses the environment, sending data to brain for processing (input).

Reading the mind

As discussed in the previous section, every participant's account of mind was strictly physical, rooted mostly in the brain, in a few cases in the body, and in one case extending beyond the

body to the physical world. With this physical understanding of the mind, it is not overly surprising that all participants believed it would someday be possible for a computer to read or decode the contents of the human mind. No participants expressed hesitation when asked about such a proposition.

For example, Alex did not feel comfortable providing a specific physical locus for the mind. Although he did not feel the probe was accurate for him, he took great pains to express his belief that such a device could work, though not necessarily by sensing the brain.

> We're driven by single-celled organisms in ways we don't really yet understand, but... there's got to be some sort of physical storage of memories or experiences. We just haven't quite learned how to read it yet. *(Alex)*

Though it leaves open room for a variety of interpretations about the exact nature of mind, Alex's view is explicit that thoughts are physical, therefore *can* be read, and *will* be read with some future technology.

There was a great deal of heterogeneity in the way this belief was bracketed or qualified. Joanna felt that there would "always be parts of the mind that can't be seen." She likened the question to the way that other people can know some parts of another person's mind, e.g. through empathy; their perspective, however, would always be partial, and she felt the same would be true for machines.

However, some participants did not bracket their belief that machines would someday read the mind. Participants for whom the authenticator worked reliably typically said that a mind-reading machine was "absolutely possible" (Mary) or "just a matter of the right data" (Alex). Participants who did not feel the authenticator was accurate described current state-of-the-art as "crude" (John) or "low-granularity" (Elizabeth).

Even Terrance, who believed the mind extended beyond the confines of the body, felt that the mind was readable by machine. After he stated his personal belief in a mind that extended to the physical environment, the researcher asked what consequence this belief might have for the future of BCIs.

> Practically, it has no implication. We could still devise an authentication tool that does the job, and it doesn't matter. Maybe in some way there could be this ESP thing where you could somehow read my thoughts... If we want to do something, we will find a way. *(Terrance)*

Terrance's language here belies broader narratives of positive technological progress (notions of "[moving] forward," and that "we will find a way"). Despite his personal beliefs about the "true" nature of the mind, he felt that engineers would manage to build the systems they intended to build, even ones with a much higher specificity than those available today (e.g. an "ESP device").

BCIs for everyone?

Generally, participants stated (implicitly or explicitly) that BCI technologies would become smaller, less expensive, more accurate, and therefore become prevalent as a consumer device. Only Mary raised the question of how institutions exert agency over the artifacts they create. Where most subjects indicated BCIs become smaller and thus more pervasive, Mary indicated that companies have beliefs, which affect what devices and technologies they produce. Specifically, Mary spoke of a "quandary" between advancing technology on one hand, and systems' autonomy on the other. She viewed this reluctance to allow systems to become more autonomous as a signal that certain technologies, potentially including BCIs, may *not* be developed for ethical, moral or philosophical reasons.

Interestingly, the other seven engineers in our study expected a future in which BCIs are pervasive, in spite of their unwillingness to wear our probe's headband in public. Some subjects believed the device's awkward, outward visibility might be mitigated by future miniaturization. Other subjects felt that social norms may simply change if the device became pervasive. This latter attitude is reminiscent of those around Google Glass, which shared an awkward (and, in practice, often stigmatizing) visibility [104]. Future work might draw out the relationship of Google Glass's imagined future to that of BCI, perhaps as a way of learning lessons about possible commercial failures, and how engineering communities may have failed to foresee them.

BCI anxieties

An important counterpoint to emerging technologies is the anxiety that rises along with them [84]. Interestingly, engineers in our study expressed no strong anxieties regarding the development of BCIs, for the most part. Regardless of their experiences with our probe, participants felt that BCIs would be developed, and would improve people's lives. Participants mentioned domains such as work, safety, and increased convenience in the home.

Only Mary reported existential anxiety about the possibility of machines that could read the human mind. She reported a technology to be "absolutely possible," and referenced the probe's continuing high accuracy as we spoke. However, in stark contrast to Terrance, Mary *feared* such a development would occur sooner rather than later.

> I hope it's fifteen years out, but realistically, it's probably more like ten. *(Mary)*

Despite Mary's prior statement about the power of institutions to change the course of technical developments, here she seems to indicate that such course changes will not occur, or that they will converge on machines that can read the mind. When pressed on downsides, the participants who did not volunteer any anxieties about BCI initially did mention security (especially the "leaking" of "thoughts") as a concern. For example, Elizabeth did not report any particular anxieties about BCIs in general, "if the proper protections are in place." Pressed on what those protections might look like, she cited encryption as a solution to privacy

concerns. Terrance, who expressed wanting BCIs to become more widespread, described in deterministic terms the cybersecurity issues such devices might pose.

> If there are security holes - which there almost certainly will be - then what happens when I'm leaking my thoughts to someone? What if I'm thinking about the seed phrase for my Bitcoin wallet... and then you put it in this anonymized dataset ... and I lose all my coins? What then? *(Terrance)*

Even alongside his concern, Terrance very much wanted a mind-reading machine to exist. He mentioned a desire for a programming assistant that would somehow speed up the process of software development. Since Terrance's conception of BCI presents high stakes with regard to privacy and security (he variously mentioned "telepathy," and an "ESP device," implying a high degree of specificity with regard to what BCIs can resolve), it is telling that he thought primarily of using BCIs to become a more efficient engineer, rather than concerns around privacy or potential harm. Later in the discussion, we unpack further how larger cultural tendencies in Silicon Valley might shape the way engineers build BCI systems.

6.5   Discussion

We find that engineers hold diverse beliefs about what the mind is, what the brain is, and about the relationship between these entities. However, all of these engineers shared a core belief that the mind is a physical entity, one that machines can and will decode given the proper equipment and algorithms. Despite this belief, engineers did not largely express concerns about privacy or security. As BCI startups continue to grow, we propose further work within technical communities, with a sensitivity toward emerging narratives, so that we may instill criticality among this emerging technical practice. We conclude with avenues for future work focusing on different communities of technical practice.

Physical mind, readable mind

Although our engineers broadly believed BCIs would become pervasive as consumer devices, we found no consistent visions of what such a future might look like. Instead, and to our surprise, we found a shared belief that there exists a physical mind that can be "read" or "decoded" by machines, despite participants' heterogeneous beliefs about its exact nature. Interestingly, only one participant shared any anxiety about this prospect with the researchers; the other participants reported looking forward to such a possibility.

Crucial to beliefs about the machine-readable mind were frames of the mind as physical, and therefore amenable to sensing. In many cases, subjects would use analogies to computation in making this point. For example, John observed an anomaly in the authenticator's performance (it did not work when he was listening to the experimenter speak). He theorized that the states are distinguishable, because speaking "is like programming" and listening to someone speak "is like being programmed". In this case, John's observations about the BCI met with

his pre-existing notions of the mind, producing a hypothesis for what "brain states" might exist *and* what states Muse headset might be able to detect. Hypotheses such as these could be consequential, as they might provide ideas or starting points for engineers looking to build systems. Our results highlight the importance of both pre-existing beliefs and particular interactions with BCIs in structuring engineers' understandings.

Broadly, engineers' beliefs about the mind-as-computer metaphor (Section 6.4) could provide starting points for engineers to build BCIs in the future. This computational view of mind has been popular among engineers at least since the "good old-fashioned AI" (GOFAI) of the 1950s. While much work has critiqued this stance from various angles [2, 48], those same critiques have acknowledged the role these metaphors have played in the development of novel technologies: If the mind is a machine, then those tools used to understand machines can also be used to understand the mind. Here, we see this metaphor return, its discursive work now focused on biosensing rather than on artificial intelligence. Of course, these metaphors illuminate certain possibilities while occluding others [48]. As such, future work should follow past research [2] in understanding what work this metaphor might do in its new domain of computational mind-reading.

Even those participants who did not subscribe to computational theories of mind still believed the mind to be strictly physical. These subjects all agreed that computers could someday read the mind, precisely because of its physical nature. While our results indicate that engineers believe the mind to be machine-readable, some work indicates that non-engineers may share this as well [4]. Future work could further investigate this claim more deeply in the context of consumer BCIs. If so, a machine designed by engineers and purported to read the mind might find acceptance among a broader public audience.

Those subjects with a computational account of mind tended to feel more confident that their account was substantially accurate. John referenced "the consensus" in justifying his beliefs about the mind being equivalent to the brain. It is worth asking whose consensus this might be: that of neuroscientists, philosophers of mind, cognitive scientists, or engineers? In any of these cases, engineers' confidence in their beliefs could have implications for what types of systems are considered buildable, and where engineers might look to validate their implementations. As products come to market, professionals in the tech industry must find ways of claiming their devices to be legitimate, or working, to the public (consumers), to potential investors, and to other engineers. These claims of legitimacy could prove to be a fruitful window for understanding the general sensemaking process around these devices as their (perceived) capabilities inevitably evolve and grow alongside changing technologies.

A future for privacy and security

Since the engineers in our study believed the mind to be readable, an important question remains around the consequences for the future of consumer privacy and security. Our participants largely acknowledged that "leaking" thoughts through security holes was a valid concern, and one participant claimed that these exploitable holes will "almost certainly" exist. However, the types of threats that engineers referenced may not square with the notion of

BCIs as a device for the masses. For example, Terrance's concern about someone stealing his Bitcoins through some BCI-based attack involves a technology which for now remains niche. This imagined scenario demonstrates how the security (and privacy) concerns of engineers may not match that of the general public. Such mismatches could have consequences for the types of systems that are designed, and whose needs these systems will account for.

Crucially, discussions about privacy and security concerns did not cause any participants to reflect further on the consequences of pervasive BCIs, nor did they deter enthusiasm for the development of these devices. These findings indicate either that engineers are not be inclined to prioritize security in the systems they build, or that they have resigned themselves to the inevitability of security holes in software. In either case, our findings suggest a long-term direction for cybersecurity concerns. These devices carry potentially serious security and privacy consequences. If our engineers will try to build devices that make judgments about the inner workings of a person's mind, future work must critically examine how to protect such systems, and the people who use them.

Implications for the design of mind-reading machines

Our findings do not indicate a singular path for the future of BCIs. Instead, they indicate an undercurrent of belief among Silicon Valley engineers in the possibility of technologies that can read the contents of the human mind. Crucially, our study revealed narratives not just around BCIs, but around the nature of the brain and mind generally, which in turn legitimize narratives about the possibility of mind-reading machines.

Despite these beliefs about what BCIs are capable of, only one participant in our study reported that ethical issues around privacy or security might deter their development. We hope engineers will become more reflexive about these beliefs around BCI, and more critical about their downstream potential for harm (e.g. surveillance). Much as utopian dialogues around the potential of the World Wide Web missed risks to privacy and security, so might similarly utopian ideals of mind-reading machines.

Since the engineers in our study believed BCIs could perform this potentially invasive "mind-reading," why did they largely want such BCIs to be built? Explanations might be found by relating the narratives we uncover to existing social and economic value systems within Silicon Valley communities. Biohacking, for one example, has become an established part of Silicon Valley culture, through dieting (e.g. Soylent, fasting), or more extreme forms of body modification (e.g. chipping) [34]. Underlying all of these cultures is a mechanical model of the body, which facilitates notions of optimization and experimentation.

How might BCIs (especially ones that purport to read thoughts) work their way into these already-established cultural patterns? We note that existing consumer BCIs already situate themselves in this context: the Muse headset we used in this study markets itself primarily as a meditation trainer (its advertising copy claims to "remove the uncertainty from meditation") [54]. Examining how BCIs perform discursive work in engineering communities will allow us to better understand engineers' intents as these devices begin to emerge, and

help us trace these intents forward as devices are re-imagined, remixed and repackaged for other groups of users in the future.

In the nascent field of consumer BCI, researchers and designers should remain in touch with the beliefs of engineers. We pinpoint beliefs about the mind, and its readability by emerging biosensing devices, as especially an critical facet. Doing so will allow design to remain preemptive rather than reactive as software for consumer BCI emerges. Designers and researchers must not remain on the sidelines; as devices come to market, we must become actively engaged in engineers' beliefs (and practices). These systems hold the potential for exploiting an unprecedented level of personal data, and therefore present real potential for harm. As such, the area presents a new locus for researchers and designers to engage critically with technical developments.

Future work

Software engineers are a diverse group, and the geographic confines of Silicon Valley do not describe all communities worldwide. Future work could explore communities in different places. Engineers in non-Western contexts may hold different cultural beliefs about the mind, which could lead to vastly different findings.

Professionals who work in machine learning could present another participant pool for future work. Machine learning is a critical component of BCIs, and many contemporary techniques, particularly deep learning, use neural metaphors to interpret and designing algorithms [6]. Thus, practitioners of these techniques may be inclined to draw metaphors between the brain and the algorithms they employ, which could color their understanding how and why BCIs work or fail.

Future work could allow participants to take an active, participatory role in the analysis of their data, and/or in the design of the BCI system. Although our participants had the technical expertise required to perform data analysis and systems engineering themselves, we did not have participants do any such analysis for this study. This participatory approach will also help us expand our understanding from engineers' beliefs to engineers' practices, as they relate to the emerging domain of consumer brain-computer interfaces. Participants might form their own interpretations of what the data mean (or can mean), building understandings that could differ from those we observed in this study.

6.6   Conclusion

As engineers in the San Francisco Bay Area, the participants in our study sit at an historical site of techno/political power. Our technology probe indicates these engineers believe the mind is physical, and therefore amenable to sensing. What are the consequences for the rest of us? I hope this study will encourage engineers to closely examine the potential of these devices for social harm, and encourage researchers to remain closely attuned to this emerging class of consumer biosensor.

What this study did not rigorously examine is how the engineers in our study encountered notions of identity as it might be captured by the brain scanning device. In general, although engineers broadly believed the mind to be readable by machines, this chapter did not deeply examine to what extent they believed the identity to be related to the mind or the brain. In the following chapter, I examine participants' responses through this lens, charting engineers' beliefs about the readability of identity as an aspect of mind.

Chapter 7

Telepathy within limits

What are the limits of machines' ability to model the mind? My arguments in this dissertation reorient this question around human beliefs: What are the limits within which claims of mind-modeling might be made (by engineers), and believed (by end-users)? I propose the term *telepathy* to describe the process of understanding models of minds. I then use this term to motivate work for charting the limits of what work telepathy might perform in the world.

7.1   Telepathy

Earlier in this dissertation, I framed prior research programs as having built *models of minds*, showing how work in philosophy supports their claims. By analyzing critiques of these research programs, I highlighted the primacy of human beliefs, both engineers' and users', in structuring how models of minds are built, and understood as relevant.

Building models of minds can be split into two major components: the engineering program of building algorithms that encode and represent mental states, and the social processes of understanding these representations as relevant in the course of life. While the boundary between these components is intrinsically unstable, the split is nonetheless useful in understanding how these models perform work in the world.

To describe the latter component, I propose the term *telepathy*. While this term has a strong connection to magic, I believe it is useful to repurpose the term for discussions about computational models of minds, and how they are understood by people. Consider telepathy's etymological pedigree in relation to other popular technologies.

Telephony (*tele + phonos*)
    Sound at a distance

Television (*tele + vīsiō*)
    Sight at a distance

Telepathy (*tele + pathos*)

Mind at a distance.

While the first two terms may have sounded like magic at some point in history, technical infrastructures have provided functionality that made these terms legible not just as technologies but as social media. *Telepathy* is in spirit no different. In relation to the other technical infrastructures, the prefix *tele-* highlights technical aspects of transmission, along with the various sociotechnical infrastructures and entanglements that make transmission, encoding, and decoding possible. Telepathy works to describe how models of minds are "made and measured" [10], while gesturing toward the unstable boundary between these two activities.

What might telepathy be used for? Answers to this question relate deeply to the beliefs of users and engineers. Thus, the relevant questions here include: What are the limits within which claims of telepathy might be made, or believed? How might emerging infrastructures of ubiquitous bodily and environmental sensing assist such claims, by ascribing higher resolution to their models? Or detract from them by making biosensory data mundane, thus challenging their presumed authority? Future work should deeply examine engineers' beliefs, how they change with evolving technologies, and how these beliefs affect (and are affected by) technical practices. Beliefs about the mind will continue to co-evolve along with our rapidly changing technical capacity to sense and model the world.

7.2   A big loop

Rather than presenting a theory of mind and a set of technologies that do or do not sense it, this work examines the relationship between beliefs about the mind and how they relate to the perceived capabilities of technology. In doing so, the cases in this dissertation gesture toward a big loop (Figure 7.1). In the right half of this loop, beliefs about the mind affect the technologies people build (and accept as working). The left half of this loop depicts existing technologies affecting beliefs about what the mind *is*.

This dissertation touched on the two halves of this loop separately, but did not speak to this loop in its entirety. This feed-forward loop between mind-reading technologies and ideas about "mindhood" raises the possibility that minds are not only readable because people believe they are, but because the very notion of mindhood will change relative to existing claims of mind-reading. How do we, through sensing minds, (re)make minds (and ourselves) through the things that sense them?

The shifting of categorical boundaries, especially as it relates to shifts in technological infrastructures, has been the concern of philosophers of technology [13] and feminist scholars [46] for many years. Future work should integrate these perspectives in an examination of the other half of our big loop, or in an examination of the loop itself. Future work could also complicate this notion of a loop, framing machines and minds as constantly co-constructed, or always entangled. An old question, "Are minds machines?" [103] could come under new light in this frame. Rather than asking what kinds of machines minds are, we may as well ask, are machine-ness and mind-ness always already entangled, and if so, what are the consequences?
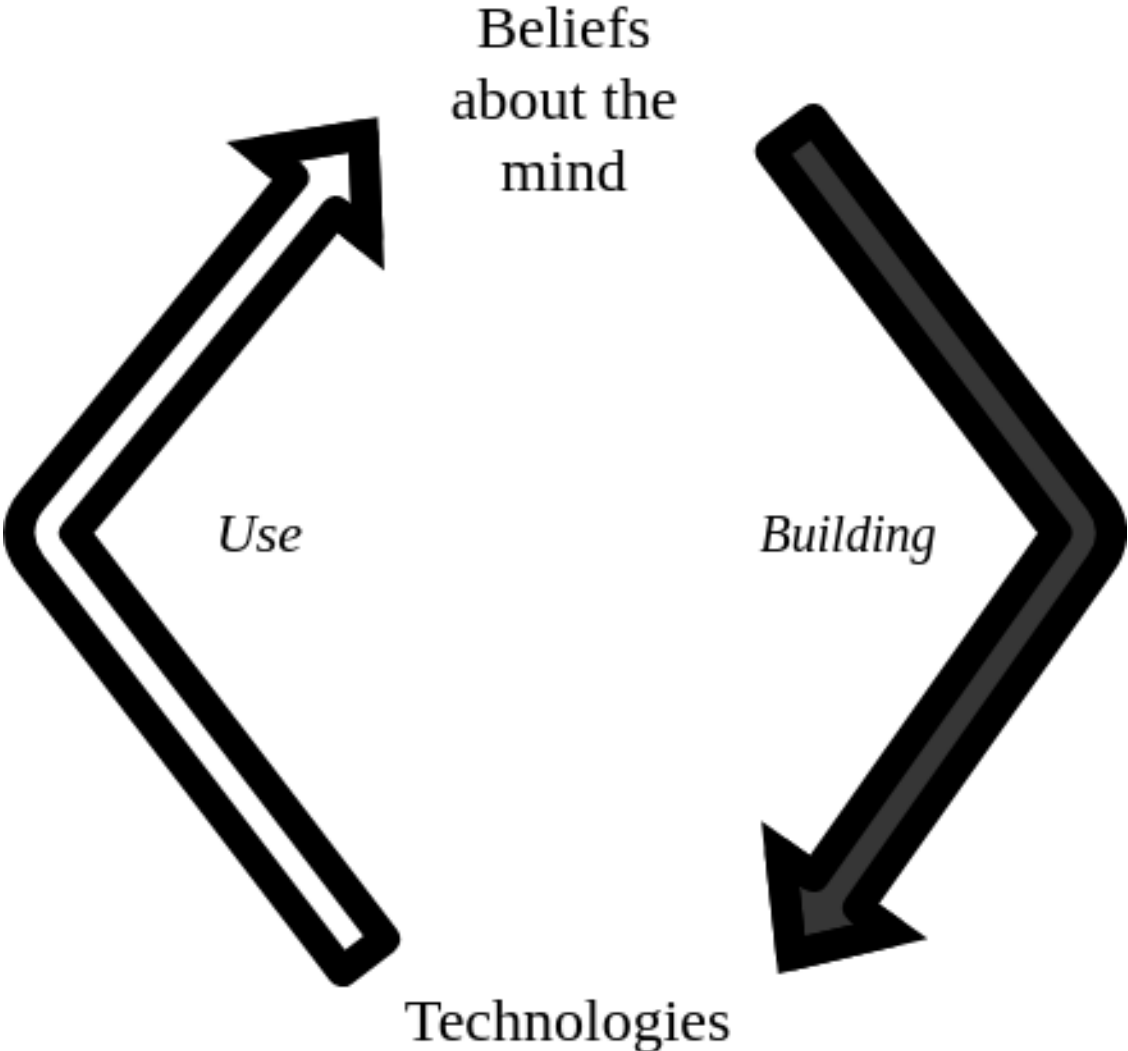
Figure 7.1: A big loop: beliefs about the mind inform the design of tools, and the use of these tools inform beliefs about the mind.

I suspect the coming years will provide opportunities to study these questions longitudinally, as technologies develop and become more diffuse. The remainder of this chapter discusses another set of longitudinal concerns, which should be studied in parallel: security, privacy, and surveillance.

7.3 Security, privacy and surveillance

While models of minds could include data about the brain, such data is not necessary to decode the mind, as this dissertation argues. Indeed, with ubiquitous enough sensing, the world at large could be (re)purposed to sense the mind. Consider the minimal example of a lightswitch. It not only takes input from people, but its design (at least the canonical version) is carefully crafted to permit only the intentional finger-action of a person. Thus, its state can be taken as a correlate of the beliefs and attitudes of the switcher(s); a request for light, a sense of darkness [98].

If IoT devices can turn anything [61] into a biosensor, what surprising features might be generated from these data? Given the potentially sensitive data that telepathy might yield, and the unclear mechanisms of intent or consent by which models of minds might be generated, future work must also engage deeply with existing work across surveillance studies, media studies and gender studies.

In Simone Browne's seminal history of surveillance in the United States [14], a racial, gendered and historical situatedness illuminates relationships between surveillance and power. While Browne's history does not paint an optimistic picture for information technologies, Mcmillian Cottom's work on black cyberfeminism [29] shows how the same tools of Browne's surveillance can be repurposed to evade surveillance, and for activism. Future work in telepathy should substantially engage with analyses such as these, so that we may better understand both what new power structures telepathy might create, and which existing ones it might (re)inforce.

Related to the sensitivity of mental data, telepathy pushes against the limits of what information assurance (IA) might mean. Traditionally, IA is concerned with the integrity, availability, authenticity, confidentiality and non-repudiation (inability to challenge authorship) of data; if the contents of mind become the stuff of data, then telepathy will plot fresh territory for cybersecurity research.

7.4 Conclusion

This dissertation aims to paint a few provocative dots on a very large canvas. As sensors continue to saturate our environment, people will continue to build increasingly high-resolution models of our bodies and minds. Machines' purported ability to divine not just what these bodies do, but what they think and feel, will prove to be a key concern for privacy, personal autonomy, and cybersecurity in the coming hundred years. It will also generate novel

opportunities for communication, accessibility, business, and entertainment. These concerns and opportunities will likely exist not in opposition to each other, but in mutual re-inforcement, entanglement, co-construction. By paying close attention to the beliefs and practices of engineers, and the expectations of end-users, we can better anticipate how (and why) the development of these technologies may occur, and thus better prepare for an increasingly connected—and increasingly hackable—world, body, and mind.

Bibliography

[1]   Affectiva Inc. *Emotion as a Service - Affectiva.* 2016. URL: http://www.affectiva.
com/solutions/emotion-as-a-service/.

[2]   Phil Agre. *Computation and Human Experience.* 1997. ISBN: 9780521384322.

[3]   Nadav Aharony et al. "The social fMRI:Investigating and shaping social mechanisms in
the real world". en. In: *Proceedings of the 13th international conference on Ubiquitous
computing - UbiComp '11* (2011), p. 445. ISSN: 15741192. DOI: 10.1145/2030112.
2030171. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-
80054080566{\&}partnerID=tZOtx3y1.

[4]   Sabrina S Ali, Michael Lifshitz, and Amir Raz. "Empirical neuroenchantment: from
reading minds to thinking critically." In: *Frontiers in human neuroscience* 8.May
(2014), p. 357. ISSN: 1662-5161. DOI: 10.3389/fnhum.2014.00357. URL: http:
//www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4034606{\&}tool=
pmcentrez{\&}rendertype=abstract.

[5]   Morgan G. Ames. "Charismatic Technology". In: *Proceedings of the 5th Decennial
AARHUS Conference* (2015), pp. 109–120. ISSN: 1944-7019. URL: http://morganya.
org/research/Ames-charisma-aarhus.pdf.

[6]   Jimmy Ba et al. "Using fast weights to attend to the recent past". In: *Advances In
Neural Information Processing Systems.* 2016, pp. 4331–4339.

[7]   SE Baron-Cohen. "Understanding Other Minds: Perspectives From Autism". In: *Neu-
rology* 45.1 (1995), pp. 210–210. ISSN: 0028-3878. DOI: 10.1212/WNL.45.1.210-b.
URL: http://psycnet.apa.org/psycinfo/1993-98373-000.

[8]   Genevieve Bell et al. "Intimate (Ubiquitous) computing". In: *Proceedings of. UbiComp
2003 . . .* Press, 2003, pp. 3–6. ISBN: 978-1-4244-5054-1. DOI: 10.1109/SOFA.2009.
5254841. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.
109.4953{\&}rep=rep1{\&}type=pdf.

[9]   Tony Bergstrom and Karrie Karahalios. "Distorting social feedback in visualizations
of conversation". In: *Proceedings of the Annual Hawaii International Conference on
System Sciences.* IEEE, 2011, pp. 533–542. ISBN: 9780769545257. DOI: 10.1109/
HICSS.2012.222. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.
htm?arnumber=6148670.

[10] Kirsten Boehner et al. "How emotion is made and measured". In: *International Journal of Human Computer Studies* 65.4 (2007), pp. 275–291. ISSN: 10715819. DOI: `10.1016/j.ijhcs.2006.11.016`.

[11] Kirsten Boehner et al. "How HCI interprets the probes". In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* (2007), p. 1077. ISSN: 1595935932. DOI: `10.1145/1240624.1240789`. URL: `http://portal.acm.org/citation.cfm?doid=1240624.1240789`.

[12] Andrey Bogomolov, Bruno Lepri, and Fabio Pianesi. "Happiness recognition from mobile phone data". In: *Proceedings - SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013*. IEEE, 2013, pp. 790–795. ISBN: 9780769551371. DOI: `10.1109/SocialCom.2013.118`. URL: `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6693415http://ieeexplore.ieee.org/ielx7/6693161/6693296/06693415.pdf?tp={\&}arnumber=6693415{\&}isnumber=6693296http://ieeexplore.ieee.org/xpls/abs{\_}all.jsp?arnumber=6693415`.

[13] Geoffrey C Bowker and Susan Leigh Star. *Sorting things out: Classification and its consequences*. MIT press, 2000.

[14] Simone Browne. *Dark Matters: On the Surveillance of Blackness*. Duke University Press, 2015. ISBN: 9780822375302. URL: `https://books.google.com/books?id=snmJCgAAQBAJ`.

[15] Winslow Burleson. "Predicting Creativity in the Wild: Experience Sample and Sociometric Modeling of Teams". In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, 2012, pp. 1203–1212. ISBN: 9781450310864. DOI: `10.1145/2145204.2145386`.

[16] Paul Campos. *The Obesity Myth: Why Our Obsessions with Weight is Hazardous to Our Health*. Book, Whole. Penguin, 2004. ISBN: 0670042846, 9780670042845.

[17] Luca Canzian and Mirco Musolesi. "Trajectories of depression". In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UBICOMP '15)*. 2015, pp. 1293–1304. ISBN: 9781450335744. DOI: `10.1145/2750858.2805845`.

[18] *Cardiogram: what's your heart telling you?* URL: `http://www.cardiogr.am/` (visited on 05/26/2016).

[19] Francesco Carrino et al. "A self-paced BCI system to control an electric wheelchair: Evaluation of a commercial, low-cost EEG device". In: *2012 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living, BRC 2012*. 2012, pp. 1–6. ISBN: 9781467324762. DOI: `10.1109/BRC.2012.6222185`. URL: `http://ieeexplore.ieee.org/xpl/login.jsp?tp={\&}arnumber=6222185{\&}url=http{\%}3A{\%}2F{\%}2Fieeexplore.ieee.org{\%}2Fxpls{\%}2Fabs{\_}all.jsp{\%}3Farnumber{\%}3D6222185`.

[20] David John Chalmers. "The Conscious Mind: In Search of a Fundamental Theory". In: *International Journal of Quantum Chemistry* January 1998 (1998), pp. xvii, 414. ISSN: 00346632. DOI: DOI10.1207/s15327809jls0803&4_10. arXiv: 0402594v3 [arXiv:cond-mat].

[21] Tianqi Chen and Carlos Guestrin. "XGBoost : Reliable Large-scale Tree Boosting System". In: *arXiv* (2016), pp. 1–6. ISSN: 0146-4833. DOI: 10.1145/2939672.2939785. arXiv: 1603.02754.

[22] C. Cheshire, a. Gerbasi, and K. S. Cook. "Trust and Transitions in Modes of Exchange". en. In: *Social Psychology Quarterly* 73.2 (2010), pp. 176–195. ISSN: 0190-2725. DOI: 10.1177/0190272509359615. URL: http://spq.sagepub.com/cgi/doi/10.1177/0190272509359615.

[23] Eun Kyoung Choe et al. "Living in a Glass House : A Survey of Private Moments in the Home". In: *UbiComp 2011* (2011), pp. 41–44. DOI: 10.1145/2030112.2030118. URL: http://dl.acm.org/citation.cfm?doid=2030112.2030118.

[24] John Chuang et al. "I think, therefore I am: Usability and security of authentication using brainwaves". In: *International Conference on Financial Cryptography and Data Security*. 2013, pp. 1–16. ISBN: 9783642413193. DOI: 10.1007/978-3-642-41320-9_1. URL: http://link.springer.com/chapter/10.1007/978-3-642-41320-9{_}1http://link.springer.com/content/pdf/10.1007{\%}2F978-3-642-41320-9{_}1.pdf.

[25] Andy Clark. "Whatever next? Predictive brains, situated agents, and the future of cognitive science." In: *The Behavioral and brain sciences* 36.3 (2013), pp. 181–204. ISSN: 1469-1825. DOI: 10.1017/S0140525X12000477. arXiv: 0140-525X. URL: http://www.ncbi.nlm.nih.gov/pubmed/23663408.

[26] Andy Clark and David Chalmers. "The Extended Mind". In: *Analysis* 58.1 (1998), pp. 7–19. ISSN: 0003-2638. DOI: 10.1111/1467-8284.00096. arXiv: arXiv:1011.1669v3. URL: http://www.blackwell-synergy.com/links/doi/10.1111{\%}2F1467-8284.00096.

[27] Sunny Consolvo et al. "Location disclosure to social relations". en. In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*. ACM Press, 2005, p. 81. ISBN: 1581139985. DOI: 10.1145/1054972.1054985. URL: http://portal.acm.org/citation.cfm?doid=1054972.1054985.

[28] K. S. Cook et al. "Trust Building via Risk Taking: A Cross-Societal Experiment". en. In: *Social Psychology Quarterly* 68.2 (2005), pp. 121–142. ISSN: 0190-2725. DOI: 10.1177/019027250506800202. URL: http://spq.sagepub.com/cgi/doi/10.1177/019027250506800202.

[29] Mcmillan Cottom. "Black CyberFeminism: Ways Forward for Intersectionality and Digital Sociology". In: (2016).

[30] Sophie Day and Celia Lury. "Biosensing: Tracking Persons". In: *Quantified: Biosensing Technologies in Everyday Life* (2016), p. 43.

[31] Michael D. Decaria, Stewart Proctor, and Thomas E. Malloy. "The effect of false heart rate feedback on self-reports of anxiety and on actual heart rate". en. In: *Behaviour Research and Therapy* 12.3 (1974), pp. 251–253. ISSN: 00057967. DOI: 10.1016/0005-7967(74)90122-3. URL: http://linkinghub.elsevier.com/retrieve/pii/0005796774901223.

[32] Laura Devendorf et al. ""I don't want to wear a screen": Probing Perceptions of and Possibilities for Dynamic Displays on Clothing". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. 2016, pp. 6028–6039. ISBN: 9781450333627. DOI: 10.1145/2858036.2858192. URL: http://artfordorks.com/pubs/16{_}CHI{_}Ebb.pdf.

[33] Cory Doctrow. *Rachel Kalmar's datapunk quantified self sensor array 2, Institute for the Future, Palo Alto, California, USA*. 2014. URL: https://www.flickr.com/photos/doctorow/15659135172 (visited on 01/31/2018).

[34] Markéta Dolejšová and Denisa Kera. "Soylent Diet Self-Experimentation: Design Challenges in Extreme Citizen Science Projects". In: (2017), pp. 2112–2123. DOI: 10.1145/2998181.2998365.

[35] Joseph Dumit. "Picturing Personhood: Brain Scans and Biomedical Identity". In: *Information Series* (2004), p. 251. ISSN: 0162-2439. DOI: 10.1353/bhm.2005.0063.

[36] Chris Elsden et al. "ResViz: Politics and design issues in visualizing academic metrics". In: *Conference on Human Factors in Computing Systems - Proceedings*. 2016, pp. 5015–5027. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858181.

[37] Deborah Estrin and Ida Sim. "Health care delivery. Open mHealth architecture: an engine for health care innovation." In: *PLoS Medicine* 10.2 (2010), e10011395. ISSN: 0036-8075. DOI: 10.1126/science.1196187.

[38] Feel.co. *Feel: The world's first emotion sensor & well-being advisor*. 2018. URL: https://www.myfeel.co/ (visited on 01/01/2018).

[39] Maridel A. Fredericksen et al. "Three-dimensional visualization and a deep-learning model reveal complex fungal parasite networks in behaviorally manipulated ants". In: *Proceedings of the National Academy of Sciences* (2017), p. 201711673. ISSN: 0027-8424. DOI: 10.1073/pnas.1711673114. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1711673114.

[40] Jérémy Frey. "Remote Heart Rate Sensing and Projection to Renew Traditional Board Games and Foster Social Interactions". In: *Proceedings of the extended abstracts of the 34th annual ACM conference on Human factors in computing systems - CHI EA '16*. ACM, 2016, pp. 1865–1871. ISBN: 9781450340823. DOI: 10.1145/2851581.2892391. arXiv: 1602.08358. URL: http://arxiv.org/abs/1602.08358{\%}5Cnhttp://dx.doi.org/10.1145/2851581.2892391.

[41] Maurizio Garbarino et al. "Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition". In: *Proceedings of the 2014 4th International Conference on Wireless Mobile Communication and Healthcare - "Transforming Healthcare Through Innovations in Mobile and Wireless Technologies", MOBIHEALTH 2014.* 2015, pp. 39–42. ISBN: 9781631900143. DOI: 10.1109/MOBIHEALTH.2014.7015904.

[42] Erving Goffman. *The Presentation of Self in Everyday Life.* Harmondsworth, 1959.

[43] Charles Goodwin. "Professional Vision". In: *American Anthropologist* 96.3 (1994), pp. 606–633. ISSN: 0002-7294. DOI: 10.1525/aa.1994.96.3.02a00100. arXiv: arXiv:1011.1669v3.

[44] Mick Grierson and Chris Kiefer. "Better brain interfacing for the masses". en. In: *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11.* CHI EA '11. New York, NY, USA: ACM Press, 2011, p. 1681. ISBN: 9781450302685. DOI: 10.1145/1979742.1979828. URL: http://dl.acm.org/citation.cfm?id=1979828{\%}5Cnhttp://portal.acm.org/citation.cfm?doid=1979742.1979828.

[45] Jun Gu, Chen-Bo Zhong, and Elizabeth Page-Gould. "Listen to Your Heart: When False Somatic Feedback Shapes Moral Behavior." en. In: *Journal of Experimental Psychology: General* 142.2 (2012), pp. 307–312. ISSN: 0096-3445. DOI: 10.1037/a0029549. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0029549.

[46] Donna Jeanne Haraway. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective". In: *Feminist Studies* 14.3 (1988), pp. 575–599. ISSN: 00463663. DOI: 10.2307/3178066. arXiv: 9809069v1 [arXiv:gr-qc]. URL: http://www.staff.amu.edu.pl/{~}ewa/Haraway,SituatedKnowledges.pdf{\%}5Cnhttps://faculty.washington.edu/pembina/all{_}articles/Haraway1988.pdf.

[47] Mariam Hassib et al. "HeartChat: Heart Rate Augmented Mobile Messaging to Support Empathy and Awareness". In: ().

[48] N. Katherine Hayles. "Contesting for the Body of Information: The Macy Conferences on Cybernetics". In: *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics.* 1999, pp. 50–83. ISBN: 0226321452.

[49] Jennifer Healey. "Physiological Sensing of Emotion". In: *The Oxford Handbook of Affective Computing* October (2014), p. 204. DOI: 10.1093/oxfordhb/9780199942237.013.023. URL: http://ict.usc.edu/{~}gratch/CSCI534/ACII-Handbook-Physiology.pdf.

[50] Noura Howell et al. "Biosignals as Social Cues : Ambiguity and Emotional Interpretation in Social Displays of Skin Conductance". In: DIS '16. ACM, 2016. ISBN: 9781450340311.

[51] Edwin Hutchins. *Cognition in the Wild.* MIT press, 1995.

[52] Edwin Hutchins. "Distributed cognition". In: *Cognition, Technology & Work* 7.1 (2005), pp. 5–5. ISSN: 1435-5558. DOI: 10.1007/s10111-004-0172-0. URL: http://www.slis.indiana.edu/faculty/yrogers/dist{_}cog/{\%}5Cnhttp://link.springer.com/10.1007/s10111-004-0172-0.

[53] Hilary Hutchinson et al. "Technology probes: inspiring design for and with families". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)* 5 (2003), pp. 17–24. ISSN: 09501991. DOI: 10.1145/642611.642616. URL: http://dl.acm.org/citation.cfm?id=642616.

[54] Interaxon. *Muse: The brain sensing headband.* 2017. URL: http://www.choosemuse.com/.

[55] Katherine Isbister et al. "The Sensual Evaluation Instrument: Developing an Affective Evaluation Tool". In: *CHI 2006 Proceedings.* 2006, pp. 1163–1172. ISBN: 1595931783. DOI: 10.1145/1124772.1124946.

[56] Joris H. Janssen et al. "Intimate heartbeats: Opportunities for affective communication technology". In: *IEEE Transactions on Affective Computing* 1.2 (2010), pp. 72–80. ISSN: 19493045. DOI: 10.1109/T-AFFC.2010.13. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5611482.

[57] Jacob Kastrenakes. *Apple Watch uses four sensors to detect your pulse.* 2014. URL: http://www.theverge.com/2014/9/9/6126991/apple-watch-four-back-sensors-detect-activity.

[58] K.N. Kay et al. "Identifying natural images from human brain activity". en. In: *Nature* 452.7185 (2008), pp. 352–355. ISSN: 0028-0836. DOI: 10.1038/nature06713. Identifying. arXiv: NIHMS150003. URL: http://www.dam.brown.edu/people/elie/NEUR{_}1680/Gallant{_}Nature{_}2008.pdf.

[59] Taemie Kim et al. "Meeting mediator". In: *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08.* ACM, 2008, p. 457. ISBN: 9781605580074. DOI: 10.1145/1460563.1460636. URL: http://portal.acm.org/citation.cfm?doid=1460563.1460636.

[60] Olave E. Krigolson et al. "Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research". In: *Frontiers in Neuroscience* 11.MAR (2017). ISSN: 1662453X. DOI: 10.3389/fnins.2017.00109.

[61] Gierad Laput, Yang Zhang, and Chris Harrison. "Synthetic Sensors : Towards General-Purpose Sensing". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2017), pp. 3986–3999.

[62] Antti Latvala et al. "A longitudinal study of resting heart rate and violent criminality in more than 700000 men". en. In: *JAMA Psychiatry* 72.10 (2015), pp. 971–978. ISSN: 2168622X. DOI: 10.1001/jamapsychiatry.2015.1165. URL: http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/jamapsychiatry.2015.1165.

[63] Lucian Leahu and Phoebe Sengers. "Freaky". en. In: *Proceedings of the 2014 conference on Designing interactive systems - DIS '14*. ACM Press, 2014, pp. 607–616. ISBN: 9781450329026. DOI: 10.1145/2598510.2600879. URL: http://dl.acm.org/citation.cfm?doid=2598510.2600879.

[64] Steven Levy. *Brain-Machine Interface Isn't Sci-Fi Anymore*. 2017. URL: https://www.wired.com/story/brain-machine-interface-isnt-sci-fi-anymore/.

[65] Andrew Liptak. "There Are Some Super Shady Things in Oculus Rift's Terms of Service". In: *Gizmodo* (2016). URL: http://gizmodo.com/there-are-some-super-shady-things-in-oculus-rifts-terms-1768678169.

[66] Gilad Lotan and Christian Croft. "imPulse". en. In: *CHI '07 extended abstracts on Human factors in computing systems - CHI '07* (2007), p. 1983. DOI: 10.1145/1240866.1240936. URL: http://portal.acm.org/citation.cfm?doid=1240866.1240936.

[67] Joanne McNell. "Who Sexts Thumbprints?" In: (2015). URL: https://medium.com/message/who-sexts-thumbprints-2138641c98c.

[68] Nick Merrill and Coye Cheshire. "Habits of the Heart(rate): Social Interpretation of Biosignals in Two Interaction Contexts". In: *Proceedings of the 2016 conference on Groupwork (GROUP '16)*. 2016.

[69] Nick Merrill and Coye Cheshire. "Trust Your Heart: Assessing Cooperation and Trust with Biosignals in Computer-Mediated Interactions". In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work (CSCW '17)*. Portland, OR, 2017.

[70] Nick Merrill and John Chuang. "From Scanning Brains to Reading Minds: Talking to Engineers about Brain-Computer Interface". In: *Proceedings of the 2018 ACM Conference on Computer Human Interaction (CHI '18)*. Montreal, QC, 2018.

[71] Nick Merrill, Max T Curran, and John Chuang. "Is the Future of Authenticity All In Our Heads? Moving Passthoughts from the Lab to the World". In: *New Security Paradigms Workshop (NSPW '17)* (2017).

[72] Cade Metz. "Facebook's Race to Link Your Brain to a Computer Might Be Unwinnable". In: *Wired* (2017). URL: https://www.wired.com/2017/04/facebooks-race-link-brain-computer-might-unwinnable/.

[73] Marvin Minsky and Seymour Papert. "Perceptrons". In: (1969).

[74] Dawn Nafus. *Quantified: Biosensing technologies in everyday life*. MIT Press, 2016.

[75] Neurable. *Neurable: Power Brain-Controlled Virtual Reality*. 2017. URL: http://www.neurable.com/.

[76] NeurotechX. *The international neurotechnology network*. URL: http://neurotechx.com/ (visited on 12/20/2017).

[77] Alva Noë and Evan Thompson. "Are There Neural Correlates of Consciousness?" In: *Journal of Consciousness Studies* 11.1 (2004), pp. 3–28. ISSN: 13558250. URL: `http://www.redi-bw.de/db/ebsco.php/search.ebscohost.com/login.aspx?direct=true{\&}db=psyh{\&}AN=2004-11344-001{\&}site=ehost-live{\%}5Cnevant@yorku.ca{\%}5Cnnoe@socrates.berkeley.edu`.

[78] Geoff Nunberg. "Feeling Watched? 'God View' Is Geoff Nunberg's Word Of The Year". In: *All Tech Considered, NPR* (2014). URL: `https://www.npr.org/sections/alltechconsidered/2014/12/10/369740829/forget-creepy-nunbergs-word-of-the-year-is-bigger-and-two-god-view`.

[79] Daniel Olguín Olguín et al. "Sensible organizations: Technology and methodology for automatically measuring organizational behavior". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 39.1 (2009), pp. 43–55. ISSN: 10834419. DOI: `10.1109/TSMCB.2008.2006638`.

[80] J. Kevin O'Regan and Alva Noë. "A sensorimotor account of vision and visual consciousness". In: *Behavioral and Brain Sciences* 24.05 (2001), pp. 939–973. ISSN: 0140-525X. DOI: `10.1017/S0140525X01000115`. URL: `http://www.journals.cambridge.org/abstract{\_}S0140525X01000115`.

[81] B Parkinson. "Emotional effects of false autonomic feedback." en. In: *Psychological bulletin* 98.3 (1985), pp. 471–494. ISSN: 0033-2909. DOI: `10.1037/0033-2909.98.3.471`. URL: `http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.98.3.471`.

[82] Brian Parkinson. "Emotions in Interpersonal Life". In: *The Oxford handbook of affective computing.* October (2014), pp. 68–83. DOI: `10.1093/oxfordhb/9780199942237.013.004`. URL: `http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199942237.001.0001/oxfordhb-9780199942237-e-004`.

[83] R. W. Picard and J. Healey. "Affective wearables". en. In: *Personal and Ubiquitous Computing* 1.4 (1997), pp. 231–240. ISSN: 16174909. DOI: `10.1007/BF01682026`. arXiv: `arXiv:1411.1164v1`. URL: `http://link.springer.com/10.1007/BF01682026`.

[84] James Pierce. "Dark Clouds , Io $ #! + , and ? [Crystal Ball Emoji]: Projecting Network Anxieties with Alternative Design Metaphors". In: (2017), pp. 1383–1393.

[85] Emilee Rader and Janine Slaker. "The Importance of Visibility for Folk Theories of Sensor Data". In: *Symposium on Usable Privacy and Security (SOUPS) 2017.* Soups. 2017. ISBN: 9781931971393.

[86] Terry Regier and Paul Kay. *Language, thought, and color: Whorf was half right.* 2009. DOI: `10.1016/j.tics.2009.07.001`. URL: `http://www.sciencedirect.com/science/article/pii/S1364661309001454` (visited on 06/23/2016).

[87] Nikolas Rose. "Reading the human brain: How the mind became legible". en. In: *Body & Society* 22.2 (2016), pp. 140–177. ISSN: 1357-034X. DOI: 10.1177/1357034X15623363. URL: http://bod.sagepub.com/cgi/doi/10.1177/1357034X15623363http://bod.sagepub.com/content/22/2/140http://bod.sagepub.com/content/22/2/140.short.

[88] Matthew Sample. "Evaluating Neural Futures: Good Technoscience and the Challenge of Co-Production". PhD Dissertation. University of Washington, 2016.

[89] AnnaLee Saxenian. *Regional advantage.* Harvard University Press, 1996.

[90] Elaine Sedenberg, Richmond Y Wong, and John C.-I. Chuang. "A Window into the Soul: Biosensing in Public". In: *CoRR* abs/1702.0 (2017). arXiv: 1702.04235. URL: http://arxiv.org/abs/1702.04235.

[91] Phoebe Sengers and Bill Gaver. "Staying open to interpretation: engaging multiple meanings in design and evaluation". In: *Proceedings of the 6th conference on Designing . . .* (2006), pp. 99–108. ISSN: 1595933670. DOI: http://doi.acm.org/10.1145/1142405.1142422. URL: http://dl.acm.org/citation.cfm?id=1142422.

[92] Roger N Shepard and Jacqueline Metzler. "Mental rotation of three-dimensional objects". In: *Science* 171.3972 (1971), pp. 701–703.

[93] Petr Slovák, Joris Janssen, and Geraldine Fitzpatrick. "Understanding heart rate sharing". en. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* February (2012), p. 859. ISSN: 0022-3514. DOI: 10.1145/2207676.2208526. URL: http://dl.acm.org/citation.cfm?id=2207676.2208526.

[94] Smith Optical. *Lowdown Focus mpowered by Muse.* 2017. URL: http://smithoptics.com/us/lowdownfocus.

[95] Jaime Snyder et al. "MoodLight". en. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15.* ACM Press, 2015, pp. 143–153. ISBN: 9781450329224. DOI: 10.1145/2675133.2675191. URL: http://dl.acm.org/citation.cfm?doid=2675133.2675191.

[96] Spire, Inc. *Spire is the first wearable to track body, breath, and state of mind.* (Visited on 05/26/2016).

[97] James Stables. *The best biometric and heart rate monitoring headphones.* 2016. URL: http://www.wareable.com/headphones/best-sports-headphones (visited on 05/22/2016).

[98] Peter Tolmie et al. ""This has to be the cats": Personal Data Legibility in Networked Sensing Systems". In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16.* 2016, pp. 490–501. ISBN: 9781450335928. DOI: 10.1145/2818048.2819992. URL: http://dl.acm.org/citation.cfm?doid=2818048.2819992.

[99] Tim Urban. *Neuralink and the Brain's Magical Future*. 2017. URL: `https://waitbutwhy.com/2017/04/neuralink.html`.

[100] S Valins. "Cognitive effects of false heart-rate feedback." en. In: *Journal of personality and social psychology* 4.4 (1966), pp. 400–408. ISSN: 0022-3514. DOI: `10.1037/h0023791`. URL: `http://content.apa.org/journals/psp/4/4/400`.

[101] Robert S Weiss. *Learning from strangers: The art and method of qualitative interview studies*. Simon and Schuster, 1995.

[102] Chris Wilson and Joe Oeppen. "On reification in demography". In: *Population, projections and politics* (2003), pp. 113–129.

[103] Terry Winograd and Fernando Flores. *On understanding computers and cognition: A new foundation for design. A response to the reviews*. 1987. DOI: `10.1016/0004-3702(87)90026-9`.

[104] Richmond Y Wong and Deirdre K Mulligan. "When a Product Is Still Fictional: Anticipating and Speculating Futures through Concept Videos". In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. ACM. 2016, pp. 121–133.

[105] Lynn Wu et al. "Mining face-to-face interaction networks using socimetric badges: predicting productivity in an ITC configuation task". In: *Icis* (2008), pp. 1–19.

[106] David Young, Richard Hirschman, and Michael Clark. "Nonveridical heart rate feedback and emotional attribution". en. In: *Bulletin of the Psychonomic Society* 20.6 (1982), pp. 301–304. ISSN: 0090-5054. DOI: `10.3758/BF03330108`. URL: `http://link.springer.com/10.3758/BF03330108`.