UNIVERSITY OF CALIFORNIA SAN DIEGO

Coping with Broken Data and Environmental Contingencies in Interaction:

An Observational Study of a Digital Mental Health Trial

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

in

Communication and Cognitive Science

by

David Benjamin Berners-Lee

Committee in charge:

Professor Morana Alač, Chair
Professor Seana Coulson
Professor Stuart Geiger
Professor Jim Hollan
Professor Carol Padden

2023

The Dissertation of David Benjamin Berners-Lee is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank some of the many colleagues, friends and family who supported and contributed to the several years of work that culminates in this dissertation. First, I would like to thank, for their guidance, inspiration and support, my advisor, Morana Alač, and the rest of my committee: Seana Coulson, Stuart Geiger, Jim Hall and Carol Padden. I am also grateful to John Haviland for his contributions to my thinking about language and interaction that have made a lasting impact on my research. I would like to thank the UCSD Department of Communication, many faculty, staff and affiliates of which have offered me support over the years. In particular, Melanie Lynn, Gayle Aruta, Gerlie Naval and Andrea Celada have been supportive in my efforts to find time, space and funding to accomplish this work. I am also grateful to current and past department chairs: Brian Goldfarb and Patrick Anderson.

I would also like to thank my fellow graduate students whose insight and support has benefitted me tremendously; particular thanks are owed to Googie Daniels, Asia Feiss, Kerry Keith, Loreto Montero, Grant Leuning, Yelena Gluzman, Magdelena Donea, Jacob Hellman, Louise Hickman, Oisín Parkinson-Coombs, Felix Binder, Cameron Jones, Ana Chkhaidze, Akshita Sivakumar, Veronica Urube, Sam Gaffney, Rachel Fox, Sookyong Ko, Dorothy Howard, Keli Gabinelli and Paula Santa Rosa Siqueira Mendes.

I am grateful for the questions, comments and discussion of the following colleagues who saw earlier versions of these ideas at conferences: Luke Stark, Marion Boulicault, Klaus Hoeyer, Wayne Beach, and Wanheng Hu.

I would like to thank the undergraduate students from the seminar I taught in Fall 2022, *Metaphor and Everyday Life,* for their substantive contributions to my thinking on issues in cognition, semantics and semiotics.

I would also like to thank Jennifer Grossman for convening the writing group at the Wellesley Centers for Women at Wellesley College and to all its participants for their writing help.

I am thankful for my on-going intellectual collaboration with Michael McGillicuddy, with whom I have had the benefit of discovering and investigating some of the theoretical issues explored in the following pages.

I would also like to thank my undergraduate mentors: Ian Blaustein, Yasser Munif and Nigel Gibson.

Many thanks to my friends and family for their patience and support. Thank you: Simon Adamson, Nathan Basch, James Fraser, Clark Jacobson, Gregory Muir, Noah Star, Saie Sabet, Saeed Pirooz, Sahnaz Pirooz, Neeka Pirooz and Charlie Bauer. I would like to thank my partner, Kianna Pirooz, for her unwavering love and support. Finally, I am grateful to my parents and my sister, Alice, for everything they have done to make this thesis possible.

Chapter 2, in part, is a reproduction of material that appears in Berners-Lee, B. (2023). Reconciling healthism and techno-solutionism: An observational study of a digital mental health trial. *Sociology of Health & Illness*, 1467-9566.13683. https://doi.org/10.1111/1467-9566.13683. The dissertation author was the sole investigator and author of this paper.

Chapter 3, in part, has been submitted for publication of the material as it may appear in *Social Theory and Health*. Berners-Lee, B. . The dissertation author was the sole researcher and author of this paper.

VITA

2017 Bachelor of Science in Politics of Representation, Emerson College

2023 Doctor of Philosophy in Communication and Cognitive Science, University of California San Diego

ABSTRACT OF THE DISSERTATION

Coping with Broken Data and Environmental Contingencies in Interaction:

An Observational Study of a Digital Mental Health Trial

by

David Benjamin Berners-Lee

Doctor of Philosophy in Communication and Cognitive Science

University of California San Diego, 2023

Professor Morana Alač, Chair

In the growing field of digital mental health, sensors and wearable devices are combined with machine learning (ML) algorithms for novel inventions. This dissertation is based on observations of the development and testing of one of these systems, which uses ML algorithms to determine correlations between behaviors and mood, and offer behavior modification plans based on these correlations, for treatment of depression. Activities observed include public talks, laboratory meetings, training of new laboratory members, and guidance sessions in which participants are instructed on how to interpret ML outputs as a basis for behavioral change and how to implement behavior modification plans. The approach of the dissertation consists in

presenting the accounts in which laboratory members describe their own activities alongside observations. Observable practices of the lab, particularly interactions that invite contingencies of real environments, are variously influenced by, in conflict with and sometimes absent from formal accounts.

Descriptions of the trial tend to emphasize participants themselves as they are revealed and empowered by the data system. Inner layers or versions of the participant are theorized as laboratory members cope with inevitable breakages and gaps in the data system as an exhaustive, neutral, and portable representation of participant behavior. What seem, in formal accounts, like manifold layers of the participant, appear in observation as the results of the relations between participants, the guides that help them to understand and implement ML outputs, and the environments in which this implementation takes place.

Formal accounts from the laboratory are explained in the context of the history of talk therapy in psychiatry, a discipline in which investigations of meaning and interpretation have become increasingly marginal. This history sheds light on a fundamental incongruence in efforts to improve ML-based mental health interventions through post-hoc explanation and interpretation practices. Rather than suggest a return to interpretive talk therapies that were prominent in mid-20th-century psychiatry, the thesis argues that attending to human interactions as constitutive of the digital mental health intervention serves to better account for new forms of human labor involved and the contingencies that emerge as participants attempt to implement these systems in diverse environments.

Introduction

1. Background

The present dissertation is based on observations of a digital psychiatry trial that takes a novel, interdisciplinary approach to treatment of mild and moderate depression. The trial draws on recent developments in machine learning and wearable technology, and puts these to therapeutic use following the clinical model of cognitive behavioral therapy (CBT). To understand the trial studied here, it is essential to understand these recent changes in the healthcare technology landscape and the modern history of CBT as a treatment for depression.

The increasing prevalence of mental health issues globally has resulted in concern about scarce resources, including the labor of mental healthcare providers (Steel et al., 2014; Thyloth et al., 2016; Whiteford et al., 2013), which, in many cases, are "insufficient to provide one-on-one treatment for most people with mental health problems" (Verhagen et al., 2022). In the United States, where the present study takes place, this issue has been acknowledged by the National Institute of Mental Health (NIMH) with optimism regarding the role of digital technologies in efficiently delivering mental health treatments to close the gap. The NIMH is a proponent of what they call digital health, which describes a combination of "mobile health" technologies like smartphones and wearable devices (eg. "smart watches"), and novel insights from genetics, biology, behavioral science and social science. Digital health is proposed as a way to improve efficiency, extend the reach of mental health experts into otherwise inaccessible communities, and bring down costs through automation (National Advisory Mental Health Council, n.d.).

Former NIMH director Thomas Insel championed one digital health tool specifically: digital phenotyping (Insel, 2017). First coined by Jain and colleagues (2015), this term refers to

behavioral health interventions that make use of smartphone and wearable data like heart rate, step rate, and step count to measure symptoms and even redefine medical conditions by their behavioral markers. While some approaches to digital phenotyping aim to combine a single participant's behavioral data with large datasets (Rüppel, 2019), many instead emphasize the strength of this approach for studies with small samples sizes and even "n-of-1" trials, which refrain from comparing participants' data and instead look to how a single participant's behaviors change over time (Onnela, 2021). Digital phenotyping is ultimately distinguished by the smartphone and wearable sensor data it is based on, and the naturalistic, *in situ* nature of data collection (Baumgartner, 2021; Birk & Samuel, 2020).

These distinguishing features link digital phenotyping to a trend called the "Quantified Self" (QS), in which ordinary people use digital devices to collect data on their behavior, often without any medical or therapeutic aims. The self-trackers of the QS movement have been said to be motivated by factors including an aesthetic interest in seeing behaviors visualized, a desire to modify what users see as the objective make-up of the body (Pols et al., 2019), a general desire for control (Ajana, 2017), and also the prospect of living an experimental life and developing a "laboratory of the self" (Kristensen & Ruckenstein, 2018). While QS users are often uninterested in medical treatments that could be informed by their self-tracking data, literature on the QS movement will become relevant to the present work. The QS literature includes discussion of the individual subject and this subject's freedom and agency, raising the question of whether users are disciplined or empowered by their digital devices and the behavioral data that these devices track, visualize and use as the basis of recommendations for behavioral change.

2. Field site: the translational psychiatry lab and the digital psychiatry trial

In the translational psychiatry laboratory that is my field site (hereafter "TPL"), weekly meetings typically include updates from lab members regarding on-going studies. Some lab members work on several studies and others focus on just one. All of these studies fall into the broad domain of translational psychiatry, a field that aims to experimentally test hypotheses about the mind while simultaneously offering treatment to participants. Studies at the lab are also related under themes including mobile technology, neural oscillations and mindfulness. In public self-descriptions, the lab's approach is termed "neural engineering." This approach is said to be leveraged towards mental health indications that have substantial impact on a societal level. Avoiding a reduction of all of psychiatry to neural circuits, the lab's public self-presentation suggests that neural circuits generally function in an interplay with the rest of the individual and their community.

On January 21st, 2022, the principal investigators of the lab announced and introduced their colleagues to a new trial that aims to treat depressed patients with personalized plans. The plans will be chosen based on data that will be collected using wearable devices and the lab's own app, and then analyzed using a range of machine learning models. This new trial, we are told, will be the most intricate, demanding and sophisticated of all of the studies in the lab and therefore will demand the contributions of most lab members.

The trial is described as a personalized, evidence-based, data-driven, machine-learning based intervention for mental health in mildly and moderately depressed patients. According to the plan for the trial, each participant will progress through a series of stages. First, they will be screened for eligibility using surveys to determine if the participant is in the correct age range for the study and whether they have any disqualifying conditions or medications. They will also be screened using standard instruments that are common in psychiatric research including the

Hamilton Depression Rating Scale (HDRS) survey and the Structured Clinical Interview for DSM-V (SCID), which serve to confirm that patients fall into the target categories of mild and moderately depressed and are not categorized as having other disqualifying conditions including suicidal behaviors.

The participant will then complete "pre-" testing including electroencephalography (EEG) testing, results of which will be compared to those collected at the end of the study. EEG testing requires the participant wear a cap which is mounted with electrodes which measure the electric field generated by neural activity, called the local field potential (LFP). Because these electrical changes can be recorded with high temporal resolution, the lab uses EEG to monitor electrical changes brought about in reaction to a stimulus, called event-related potentials (ERPs), using their own app to provide cues and stimuli during EEG sessions. LFP changes occur not only in response to stimuli, but also in continuous oscillations that have been theorized as a substrate for high-level communication between different parts of the brain (Fries, 2005). These neural oscillations also will be captured in this phase of testing.

After completing the pre-screen, the participant is fitted with a "smart watch" and introduced to a mobile application. The smart watch will automatically record their heart rate and monitor their movements using on-board gyroscopes and accelerometers. The mobile application will elicit participant self-assessments of mood, exercise and diet (for screenshots of the application as participants see it, see Chapter 5). These prompts by the mobile app constitute ecological momentary assessments (EMAs).

For the next phase which lasts at least 15 days, the participant will produce data. They will use the digital instruments to collect the data but these instruments will not suggest any intervention at this phase. The goal of the phase is to produce adequate data for the "data team,"

lead by a graduate student in computer science, to develop a machine learning prediction model. A sufficient dataset consists of 15 days of watch data: 15 complete EMAs (to be completed once per day) and 45 more basic mood assessment EMAs (to be completed three times per day). The period of 15 days is not an essential component of this plan. Some EMAs can be missed, but this will result in a longer duration of this phase because a total of 60 EMA data points are required before the participant can move on to the next phase of the study.

After this initial dataset is complete, the data team will test various machine learning models, including random forest, Poisson regression and elastic net. The team will select a model for its statistical power in predicting a change in mood on the basis of another parameter that can be modified by a behavioral plan. There are four behavioral plans: diet, exercise, positivity and sleep. Plans are named for the activity that will be the target of behavioral modification, and are largely based on interventions common in cognitive behavioral therapy.

After a behavioral modification plan is selected for the participant, the intervention will begin. The participant will be guided on how to interpret the machine learning model's output and implement the proposed behavioral modification plan. Before this process can begin, however, the data team will explain the machine learning (ML) model's outputs to the clinical guides as they prepare to begin guidance. Guides will prepare for the guidance process by completing a series of trainings that simulate the guidance situation. Guidance sessions are planned to take place weekly. Sessions start with an introduction to the participant's assigned plan: sleep, diet, positivity or exercise. After familiarizing the participant with the overview of the plan and the relevant handout, subsequent guidance sessions consist of setting weekly goals and evaluating the previous week's behaviors. This process is described as "not therapy." Instead it is aligned with motivational interviewing: one goal is to direct and augment the participant's

intrinsic motivation and assist the them as they follow the ML model's guidance. If the participant lacks motivation entirely the intervention is bound to fail.

The intervention will last six weeks. At the end of these six weeks the participant will undergo "post" testing, which will be conducted using the same instruments as "pre" testing, including EEG, while using the app as well as screening questionaries (HDRS and SCID). The same measurements are taken at the close of the study to enable comparison between pre- and post-intervention.

3. Methods

The present study is based on qualitative observations of laboratory practices that have much in common with ethnographic observations. However, they differ from those conducted for a traditional ethnography due to the remote, mediated nature of everyday work at the TPL. Observational data are collected in order to understand how members of the laboratory practically accomplish their work. Particular emphasis is placed on the interactional accomplishment of the guidance sessions, in which guides support participants as they attempt to implement ML system recommendations. Observational data are analyzed using conventions from conversation analysis, which aim to reveal the contingent, local and particular factors that constitute laboratory work. This emphasis on the practical accomplishment of interactional work is supplemented with an investigation into the formal accounts, schematics and conceptual systems through which laboratory members describe their own activities.

Both the methods and the research questions for this dissertation are oriented around the notion that formal accounts by which scientists conceptualize their activities inevitably affect the way these scientists carry out their work as it appears to a social scientist in observation. And yet these conceptualizations are also partial, and inevitably diverge from the observed practice. Discourse analytic methods are therefore leveraged to analyze formal accounts by which lab

members describe their work, and then these accounts are put beside observational data to reveal the connections, gaps and disjunctions between the two.

3.1 Observational data collection protocol

Before observing any laboratory member or participant I send them my IRB information sheet and go over its key points, including a basic description of my project aims and assurance that they are welcome to request that I stop observing at any time and may also ask that I delete any record of any past interaction at any time. Lab meetings, data team meetings, guide trainings and participant guidance meetings are the most common settings of laboratory interaction, and all of these interactions, as is the case of all interactions involved in the trial, occur online using ZOOM web conferencing software. After informing participants, I record these ZOOM meetings on my local hard drive. I take notes during observations and do not disturb the natural course of interaction. After an observation session, I automatically generate a provisional transcript using Microsoft speech-to-text software. Other than video recordings, I also collect data types including datasets, word processing documents, presentation slides, pieces of the software being developed and snippets of code.

3.2 Challenges to the tradition of ethnographic observation: digital, remote and trace ethnography

The remote nature of my observations of ZOOM meetings, which constitute much of the data analyzed for this dissertation, raises questions about the prospect of applying ethnographic methods to a trial that was itself conducted mostly over teleconference. Digitally mediated interactions are not equivalent to those that occur in person. The remote observer does not occupy the shared space in which participants position themselves, share artifacts, and gesture to one another, for example, which have been shown by scholars in science and technology studies (STS) as rich components of interactions that constitute scientific work (Alac, 2011; Becvar et al., 2005; Leyland, 2016; Ribes et al., 2013).

In this case however, the limitations of remote, digitally-mediated interactions apply just as much to the work being observed as to the observation process. Therefore, the reflexivity common in the ethnographic method is, in this case, directly relevant to the observed practices (Beaulieu, 2017; Bemme et al., 2020). The observing social scientist and the laboratory trainer, trainee and participant all face the challenge of engaging, listening and participating with remote interlocutors, and all benefit from the unique interactional affordances of the teleconferencing situation (Grosjean et al., 2021). The reflexivity that distinguishes ethnographic methods accords with the principle of "unique adequacy" in ethnomethodology, which holds that the phenomena of everyday practical activities always include within them adequate methods for observing them as they are produced locally (Garfinkel & Wieder, 1992).

Beyond the lack of a shared interactional space, the trial also differs from the scientific work observed in traditional laboratory studies in its highly scheduled nature. Collaborative work at the TPL takes place largely in scheduled meetings. The ZOOM meeting schedule imposes a structure onto the trial that does not exist in spatially co-present scientific collaboration, in which tasks flow into one another and sometimes occur spontaneously.

Rather than simply a lack of embodied co-presence, remote, digital collaboration and communication is analyzed here as itself a complete interactional context, leading to questions about how it is possible to communicate about embodiment, affect, mood and the experience of depression, without the interactional resources of co-present gesture and use of physically shared space and artifacts. This is particularly relevant given the broad societal context of the COVID-19 pandemic, during which scientists and ethnographers alike were forced to suspend in-person interactions.

Among the interactional artifacts that become especially relevant in the remote, digital work context are technical representations, including snippets of code, machine learning outputs in the form of tables and figures, spreadsheets that log participant meetings and plans and other forms of documentation that Geiger and Ribes (2011) call "traces." Following these authors' "trace ethnography" approach, I attend to the technical documentation that forms the digital context of laboratory interactions. I take particular interest in how these representations are foregrounded and hidden through in-meeting affordances like screensharing, and through aspects of the broader technical architecture of the trial which determines, for example, which figures participants have permission to access. Technical traces are analyzed here to understand how laboratory members accomplish the day-to-day work of running the trial, rather than as their own aesthetic or ideological whole.

3.3 Formal accounts

The PI of the lab presented the design of the TPL trial at conferences, promoted it on television, and repeatedly explained it, in its various iterations, to the rest of the lab. This presentation, which is one example of what is here called a formal account, cannot be presumed to have any particular relationship to the practices that will be carried out as the project gets underway. It does, however, show how the PI conceptualizes the project. Also, public presentations demonstrate what, in the view of the PI, is at stake for the public. Michel Callon uses the notions of framing and overflow to describe this distinction: framings are the abstractions we use to separate entities from complex networks; overflows are the contradictions, problems and novelties that draw us back to the partial nature of the our framing and inspire new framings (Callon & Rabeharisoa, 2008; Winthereik & Langstrup, 2010). For Garfinkel, the process by which sociologists and other professionals represent activities through formal concepts is called a "rendering theorem" (Garfinkel & Wieder, 1992). While ethnographic investigations often

highlight the overlooked differences between what scientists say they do in formal accounts and what they actually do in quotidian practices, the formal accounts themselves nonetheless play an important role in the course of activity (Clayman & Maynard, 1995, p. 10).

For example, the expectations produced by the presentation of prospects in advance of a biotechnological development have been shown to influence its practical accomplishment (Brown & Michael, 2003). In digital psychiatry specifically, new advancements are often promoted through attempts at "field-building" through promissory descriptions of projects yet to get underway. Martyn Pickersgill uses the term performative nominalism to describe this ability to elicit funding and constitute a new field through the development of neologisms and novel descriptions (2019). This term builds on Ian Hacking's notion of dynamic nominalism (2002), which sees naming practices as one part of the process by which new categories become real. Observations of lab practices will supplement our inquiry into formal accounts and thus address the other component of Hacking's dynamic nominalism: how things and people in the world are practically handled, represented and produced to fit into formal concepts and categories, challenge or upend these categories, or inspire new ones (Bowker & Star, 2000).

3.4 Historical and political economic context

It may seem that attention to the particular practices of the lab, as they are carried out in particular local contexts, cedes broader issues around the historical trajectory and political consequences of digital mental health. However, to the contrary, an orientation toward the particulars of how interaction is accomplished offers a way of seeing the partial nature of formal accounts, critically reflecting on these accounts, and developing recommendations that are based on the practical context of laboratory work. In an introductory chapter that Lucy Suchman added for the second edition of her influential monograph, *Human-Machine Reconfigurations: Plans*

and Situated Actions, the author explains the importance of attending closely to interactional phenomena, especially for gaining insights into broader societal issues:

> There is in my view no inherent conflict between an ethnomethodological approach to studies of situated action and an interest in cultural historical continuities and their effects. The commitment to situated action orients us, however, always to the question of just how, and for whom, culturally and historically recognizable formations take on their relevance to the moment at hand… Tropes of "large" and "small," "top and bottom," give way to analyses of the cumulative durability and force of practices and artifacts extended through repeated citation and *in situ* reenactment. Ethnomethodology and other poststructuralist approaches to social order propose, in sum, that it is only through their everyday enactment and reiteration that institutions are reproduced and rules of conduct realized. (Suchman, 2007, p. 16)

Following this perspective, the present dissertation will leverage analyses of interaction to gain insights into issues around labor equity in digital health, the prospect of interpretation and explanation serving as safeguards against inequitable uses of AI, and other pressing sociological issues. The potential to contribute to pressing, critical debates on these topics is afforded, in part, by the strength of the present approach in exploring formal accounts as incomplete, partial schemas that have a problematic relationship to the practical course of action they represent. In his contribution to a volume on ethnomethodological approaches, Randall Collins highlights this aspect of what he calls "micro-sociology": "Actors in everyday life and sociologists in their analytical constructions are alike … in using reifications and thereby contributing to the privilege of those individuals who benefit from the deference they thereby receive in real-life micro-situations. In this critical stance, radical micro-sociology can be very radical indeed" (Collins, 1981, p. 86)

4. Stakes and gaps in the existing literature, limitations of the present study

This dissertation responds to what has been called the "gap between the development of [a] robust algorithm and its clinically meaningful application," "the AI chasm," and "the last mile from better data to better care" (Insel, 2017; Keane & Topol, 2018; McCradden et al., 2022). These expressions describe the challenges of implementation, communication and interpretation that emerge as machine learning algorithms are used for clinical applications. These aspects, compared to the technical challenges of developing algorithms, machine learning based software and the hardware needed to collect input data, is a blind spot in digital psychiatry research. Specifically, the present study focuses on what Kelly et al call "human barriers:" "even with a highly effective algorithm…it will be important to maintain a focus on clinical applicability and patient outcomes, advance methods for algorithmic interpretability, and achieve a better understanding of human–computer interactions" (2019, p. 5). Along these lines, I focus on the interactions through which developers, clinicians and participants make sense of machine learning outputs, the new forms of nontraditional staff required to reconcile diverging interpretations, and the environmental factors that complicate participants' efforts to enact machine learning based recommendations in the real world.

Enhanced interpretation and explanation are frequently recommended avenues for addressing bias in machine learning as it is applied in digital phenotyping, including by Chandler et al., who propose "explanability" as a way to hold relevant components of digital mental health systems accountable for failures and biases (Chandler et al., 2019). Participants' individual interpretations of machine learning outputs and how they relate to their health, mood and behaviors may stymy the effectiveness of algorithms which may have been proven efficacious in the lab. Existing inquiries into bias and explanation in machine learning are not adequate for assessing digital phenotyping interventions like the one studied here. An influential report by

Chen et al, as well as work in the subfield called "eXplainable Machine Learning" (XML) from Rasheed et al, both focus on bias in machine learning in healthcare generally, basing their analysis on problems that emerge from large, anonymized training datasets collected through traditional clinical methods, rather than the n-of-1 approach and wearable data collection devices used in the trial we observe and many other digital phenotyping programs. Because of the novelty of the digital phenotyping approach in both its data collection and analysis methods, the challenge of explanation is unique from those addressed by XML and other established literatures. Further, these and other analyses of machine learning in healthcare fall short of applying to digital phenotyping because, as is explained in depth in chapter 2 of this dissertation, psychiatric interventions make use of meaning and communication as a component of therapies themselves, rather than merely as a justification for a proposed therapy (Alfaras et al., 2020; Hirsch et al., 2017). Because of the centrality of meaning and interpretation in digital mental health, recommendations must leverage rich qualitative data toward inductive, interpretive analyses. The use of interdisciplinary teams also distinguishes digital phenotyping from other health applications of machine learning—the different disciplinary trainings of TPL members, as will be discussed in chapter 3, makes differences in conceptual and interpretative systems more likely to occur and more profound when they do (Koutsouleris et al., 2022, p. 834; McCradden et al., 2022, p. 8).

While nontraditional mental health professionals have been pointed out as a potential confounding variable (Baños et al., 2022; Torous et al., 2018) and as a common feature of trials that achieve larger effect sizes (Linardon et al., 2019), their work practices have yet to be rigorously specified. Ethnographic studies in the social sciences have attended to related instances of care work that bridges gaps in interventions that are based around the

implementation of new technologies (Pink et al., 2018; Schwennesen, 2019; Semel, 2021). Bringing attention to overlooked forms of care work is a strength of STS approaches to interventions—especially those that are described in formal accounts as based on new technologies (Gray & Suri, 2019; Irani & Silberman, 2013). By centering interpretive and interactional work practices, my methods are well-suited for specifying new, non-traditional kind of labor involved in the digital mental health intervention.

In a recent review from *Lancet Digital Health,* Koutsouleris and colleagues explain that "studies have repeatedly shown how individual technology-based mental health interventions do not exist in a vacuum; rather, they form a part of socially situated and structurally influenced pathways to care," lamenting that "most machine learning solutions in mental health are being developed in silos, without the inclusion of decision makers, experts, and end users" (2022, p. 835). Most analyses of inequalities in machine learning based healthcare focus on traditional clinical environments, or the ways existing inequalities influence training datasets used for supervised machine learning (Chen et al., 2021). The digital phenotyping data collection context, on the other hand, demands dedicated study regarding environmental factors because of how it operates in the everyday living and working environments of participants which constitutes what Kelly et al call the "live clinical environment" (Kelly et al., 2019, p. 6; Koutsouleris et al., 2022, p. 834). As we see below in Chapter 3, this a popular topic with participants in guidance meetings and it has consequences for their ability to successfully take part in the trial.

The present study will address this issue by accounting for the environmental factors that participants report as relevant in their attempts to implement machine learning based recommendations. By basing this inquiry on participant reflections that take place in guidance meetings, it will also serve to bring participant voices into discussions of implementing machine

learning systems, while simultaneously acquiring rich qualitative insights on the difficulties posed by environmental factors like the infrastructure of home environments, the temporal requirements of various occupations and the scheduling demands of caregiving responsibilities.

Finally, the present dissertation will address methodological questions about the relationship between the methods I use for observations and analysis, and those used by members of the TPL. Of particular interest are the commonalities between components of ethnographic work, including presence and engagement, and those of guidance work.

Following the first digital phenotyping initiatives, which began with the work of Jain et al. (2015), and enthusiastic championing of these systems from former National Institute of Mental Health director Thomas Insel, sociological work has responded to the promise of digital phenotyping and its early instantiations, and also the technologies that constitute digital phenotyping systems including wearable devices, digital health apps and AI or ML based recommendations (Birk & Samuel, 2020). Common themes in this literature include articles on self-tracking with wearable devices as a form of surveillance capitalism or "participatory surveillance" (Cosgrove et al., 2020; Esmonde & Jette, 2020), the risks of privacy violations and bias involved in these applications (Christovich, 2016; Loi, 2019) and, most prominently, arguments about the reduction of complex activities and affective experiences of everyday life to quantitative metrics (Berkhout & Zaheer, 2021; Lupton, 2013b; Maslen, 2017). Another notable challenge to wearable-based digital health programs is the argument that the kinds of data that are deemed valuable in these systems are more easily produced by users of higher socioeconomic status (Charitsis, 2019; Sheller, 2018 cited in Birk & Samuel 2022).

While these contributions help offer context in terms of the political economic picture of digital mental health, none offers a practice-oriented analysis, or studies a digital health trial like

the one studied here. Other pieces of the substantial social science literature on digital mental health, many of which offer important conceptual or historical grounding for the present study, are described in each of the following chapters.

The present dissertation contributes to important debates in the social science of digital mental health, but is also limited by its methods, scope and field site. Because of the theoretical basis of its approach, this dissertation does not engage with discussions of the aesthetic or cultural value of ML-outputs or experience of participating in the trial. It also does not evaluate the trial in terms of particular political values or as praxis. This dissertation is also limited because its observational material does not include the experiences of participants' day-to-day experiences producing data in their living and working environments. This kind of material is potentially accessible through other research methods including the walk-through method (Light et al., 2018).

5. Summary of Findings

The thesis contends that the problems of explanation and interpretation in ML and the problem of implementation and environmental factors in psychiatry can be addressed together by attending to the human interactions that constitute the digital mental health intervention. In observations transcribed in Chapter 2, PWP guidance appears as a site of engagement that elicits reflections from participants and freely flowing interactions between guides and participants. In these interactions we see connections between the process of interpreting ML outputs and the contexts in which participants attempt to enact the recommendations that were made on the basis of those outputs. In chapter 5, this link between adequate explanation and coping with contextual factors continues into discussions around participant self-reports and the problem of reconciling these with passively sensed data.

Formal accounts from the TPL, through which scientists describe their own work to one another and to the public, place a limited importance on sites of interaction. The thesis demonstrates how the participant appears in these accounts as multiple or multi-layered, and how this conceptualization relates to the notion, common in the earlier psychiatric theories explored in Chapter 1, of multiple, sometimes contradictory meanings underlying an entity. The complexities of this multi-layered participant appear in observations presented in Chapter 3, not as layers of an individual, but in the dynamic relations between the individual, PWP guides, representations of behaviors and moods, and the contexts in which participants try to change behaviors. Specifically, representations of these behaviors and moods include the EMAs that participants submit using the mobile application for the trial. Contextual factors, which become bound up with these reports in interaction, include neighborhood infrastructures, work schedules and social commitments with friends and family.

The thesis also contends, in Chapter 2, that the role of PWP guides is established in formal accounts as one that is supplementary to the technical, data-driven components of the intervention. Following this idea, guides attempt to assimilate the language of machine learning into their practice, but inevitably end up also engaging in more open-ended, reflective and interpretive discussions. The transition from the guide who presents themself as a technical support to one who engages in open-ended interactions is shown as an interactional process (Chapter 5, Section 5.2). Also presented as an interactional practice, in Chapter 4, is the process through which guides are trained for a role based around adequately responding to participant queries and *saying the right thing at the right time* to resolve issues that participants might raise. Interactional resources including hypothetical active voicing emerge as tools in this training process. The discrepancy between guides and participants as they appear in formal accounts and

the way they appear in observation, itself becomes an observable phenomenon, as PIs train guides on the role as they conceptualize it and participants challenge guides by asking about the extent of the role.

These findings demonstrate the importance of including and attending to open-ended, interpretive forms of human interaction that are already involved in the digital mental health trial, rather than minimizing their importance relative to machine learning algorithms and new technologies. In these interactions, the problem of making a behavior modification meaningful, combines with discussion of contexts, like neighborhood infrastructures and work schedules. A focus on interaction can contribute to a holistic effort at improving the participant experience of implementing a digital mental health intervention. This is presented as an alternative to addressing environmental factors and explainability in ML as two separate external problems, which interfere with the substance of an intervention.

6. Overview of Chapters

As described in the above Methods section, my approach is based on analyzing both the formal accounts and conceptualizations of the trial that are developed by laboratory members and also the observed practices which may take their structure from, conflict with or otherwise relate to formal accounts. The first substantive chapter of the dissertation provides histories of relevant concepts from psychiatry, psychology and cognitive science. Each of the following substantive chapters responds to the following framing questions: How do laboratory members conceptualize the relevant practice or set of practices when they describe them before and after they take place? How does this conceptualization relate to theories from the history of the mind & brain sciences, computer science and data science? How and to what extent does this conceptualization appear

in observable, embodied practices? What observed practices escape, challenge or conflict with formal accounts and what are the consequences of these oversights and contradictions?

Chapter 1: Personalized and Data-Driven Psychiatry: History and Context

This chapter explores the history of concepts from psychiatry including the idiographic–nomothetic distinction that forms the basis of the "n-of-1" methodology, and the personalization component of the trial more broadly. Also explored in this chapter is the history of dialogic forms of talk therapy as it has been influenced by cognitivism and various forms of behaviorism. Particular attention is paid to the emphasis on intersubjective processes of communication and interpretation and ways in which these processes were progressively devalued and discarded through the rise of cognitive behavioral therapy. Finally, these histories are connected to the more recent history of explanation as a concept in computer science, artificial intelligence and machine learning. In sum, these historical discussions make it possible to establish more robust and specific research questions for later chapters and demonstrate connections and tensions in the history of the various disciplines in which TPL members are trained.

Chapter 2: Reconciling Healthism and Techno-solutionism in Formal Accounts at the Translational Psychiatry Trial

Chapter 2 consists of side-by-side analysis of formal accounts and observed practices. Formal accounts are shown to establish a conceptualization of the trial as based primarily on a participant, who has internal capacities and motivations to change behavior for the better, and the algorithmic technology, the insights of which enable the participant to access their potential. This two-sided formulation fails to adequately accommodate the role of guides, who provide support to participants as they attempt to implement ML recommendations. Observed interactions demonstrate how guides align themselves with the technical side of the formal account, taking cues from the schemas of the machine learning outputs and applying technical verbiage from machine learning to participants' activities, affects and moods. This chapter demonstrates the

importance of specifying new forms of human labor, even when they are deemed supplemental in formal accounts.

Chapter 3: Care practices and Environmental Factors Behind the Multiple Selves of the Digital Psychiatry Participant

Chapter 3 also starts out by analyzing formal accounts through which laboratory members describe their work in advance of carrying it out or in training their colleagues. The participant who is said to be empowered by insights from the machine learning algorithm is also said to have multiple versions or layers of the self. The participant is said to include, in an inner capacity, not only the motivation to change behaviors, but also the knowledge of which behaviors should be changed, and the improved mood and affect that is the goal of the intervention. This conceptualization of the participant as a multi-layered agent who, empowered by machine learning insights, is capable of bringing about their own mental health treatment, is analyzed as part of a historical shift in psychiatry away from meaning and interpretation.

Chapter 4: The Social Organization of Guidance Training

Chapter 4 focuses on how laboratory members train guides in preparation to support participants. Using conventions from conversation analysis, this chapter specifies the interactional repertoires that principal investigators (PIs) of the lab use to instruct guides. PIs use a conversational resource called hypothetical active voicing (HAV) to "quote" future speech that the guide or participant might engage in. HAV has been shown to play an important role in CBT, serving to pre-empt patient resistance that might take the form of "I don't know what I would say." Chapter 4 specifies how HAV is marked in the training context and demonstrates how pre-empts substantive discussion of sociotechnical issues with the trial, effectively reorienting the training session around how to cope with participant resistance. This move is then explained in the context of formal accounts that see the purpose of guidance sessions as transparent transmission

of ML recommendations, rather than the dialogic process of eliciting, responding to and reflecting on participants' talk, which constitutes much of the observed guidance process.

Chapter 5: Coping with Breakages and Contradictions in Self-Reported and Passively Sensed Data

The final substantive chapter explores self-reports of mood and exercise and also the "passively sensed" sensor data that is automatically gathered by their smart watches, and how these data types are discussed by laboratory members when unintuitive or contradictory findings emerge in analysis. Participants attempt to reconcile disjunctions between these two kinds of data in guidance meetings. Guides are shown to play a mediating role, interacting with PIs in one setting and participants in another. In guidance sessions and laboratory meetings, differing ways of accounting for mood and exercise are reconciled, compared and sometimes siloed for the sole consideration of either just laboratory members, or just the participant and their guide. Interactions involving two participants are shown as contrasting case studies. Both participants produce self-reports of mood and exercise that contradict passively sensed data, and in both cases laboratory PIs discuss this problem in meetings with the participants' guides. Subsequent guidance sessions show how participants attempt to defend their own accounts of their bodily experience, each participant taking a different approach. One participant bases her explanations on particularities of the environment and temporal context, and the other attempts to explain her data using rhetoric from the health sciences. Findings from this chapter demonstrate the challenges for participants and laboratory members of reconciling digital data, self-reports and participant explanations of how they complete self-reports. Specifying these processes demonstrates the challenge of consulting patients and participants as a safe-guard against harmful or misguided behavioral interventions.

Chapter 1 Personalized and Data-Driven Psychiatry: History and Context

1. Introduction

The ways that laboratory members conceptualize their work influences the course of their scientific practices including, for example, PWP plan assignment and PWP guidance. Because this dissertation attends to both these formal accounts and also observable practices, it is necessary to explore the historical and disciplinary connections between concepts like personalization, depression, thought, mood and affect. This includes looking back at the history of the disciplines of psychiatry and psychology from which contemporary mental health treatments emerged and also looking across to related contemporary domains in which these concepts are taken up for other data-driven, personalized interventions and services. This chapter will lay the groundwork for chapters 2 and 3, which rely on analysis of the theories that scientists use to describe their work. It also will demonstrate connections, insights and ironies that arise when the histories of relevant disciplines are combined with a survey of contemporary applications.

Of particular interest here are concepts that TPL members use in general descriptions of the trial. Personalization and the empowered, individual participant that personalization centers around, is the first of these concepts. This chapter explores the history of personalization as a concept in psychiatry and how it has been taken up by other contemporary ML-powered systems.

The data-driven nature of the trial is also analyzed here in the context of the history of cognitive behavioral therapy and psychoanalysis, demonstrating how the description of the TPL trial as data-driven can be seen as the result of a broader historical trajectory in the discipline.

These historical and conceptual contexts are offered to also lay the groundwork for asking the right questions when we turn to empirical material from the field site. The question of how to improve explanations of ML outputs in mental health care is shown below to be inherently challenging because of a broader disciplinary shift in psychiatry away from questions of meaning and interpretation.

2. Personalization, The Individual and the Idiographic

A popular buzzword in contemporary psychiatry and mental health discourse, *personalization* is typically presented as a feature of a proposed intervention. In digital psychiatry, it often plays a major role in leveraging the immense quantities of personal data that are continuously collected by smartphones. Discourse around individualization and personalization is an appropriate place to begin our look into formal accounts of the trial because unlike the non-human entities of the world, which are rigidly understood as real in the course of what Kuhn (1996) calls 'normal science,' "in [the case of] social phenomena we may generate kinds of people and kinds of action as we devise new classifications and categories… We 'make up people' in a stronger sense than we 'make up' the world" (Hacking, 2002, p. 40). The practical relevance of names and categories has been studied not only in the digital psychiatry context, but also in the context of machine learning and personalized apps by Robert Prey who, modifying Raymond Williams' adage that "there are in fact no masses, but only ways of seeing masses," suggests that "there are in fact no individuals but only ways of seeing people as individuals" (2018, p. 1087). Both the psychiatric concept of the idiographic intervention as well as trend of personalization in recommender systems will help to understand the individual at the heart of the TPL's

personalized intervention as the result of the sociotechnical process by which new kinds of individuals are continuously produced.[1]

Personalization is a key component of the study as it is described by the lab's principal investigator (PI). It is also emphasized in public presentations of the trial including in a local news segment. Not only is the ML system trained on a participant's own data, but the model is selected for its suitability in analyzing their dataset — the algorithm itself is selected for the individual. Data from other participants does not influence this process. The approach is called *idiographic*, or *N-of-1* (referring to the number of participants (N) involved in the development of the model). While the project originally sought to offer unique plans to each participant, this was scaled back as the study design was finalized. Instead, the finalized study design uses personalized machine learning systems to decide on one of four generic plans. In this way, it is not the final recommendation that is unique to the participant, but the manner in which the participant is measured in order to choose a plan.

2.1 The idiographic/nomothetic distinction in psychiatry

This individual, personalized approach is described by the PI as *idiographic*, a term which is commonly used to refer to psychological findings based on investigations into unique individuals as opposed to "nomothetic" investigations into traits that are shared among a class or type. The term *idiographic* has come into use as a synonym of *N-of-1*, and it has been suggested by prominent institutions, including the World Psychiatric Association (WPA), that new therapies

[1] Discussion in the field of critical data studies of the constructed individual user also often makes room for an alternative notion of an individual as defined by things outside of the digital realm (Seaver, 2015). Because of this effort to retain a humanistic individual beside the digital one, the individual produced in big data projects has been termed a "data double" (Ericson & Haggerty, 2000) or "data shadow" (Andrejevic, 2013).

should prioritize this kind of intervention.[2] The meaning of the term shifted considerably over the course of the 20th century.

The term *idiographic* was first used by German philosopher Wilhelm Windelband, who minted the term in the context of a distinction between the aims of natural science disciplines (*die Naturwissenschaften*) and those of sciences of spirit or mind (*die Geisteswissenschaften*). The former he called *nomothetic*, the latter *idiographic*. The distinction was based not on the content of the subject studied, nor the particular methods taken, but rather "the kind of knowledge sought after" (Lamiell, 1998, p. 27; Windelband, 1998, p. 12). Nomothetic investigations pursue "what always is"—general laws that are not historically contingent. These laws allow us to know specific instances in advance, as all instances are bound to follow any given law. Idiographic ones, on the other hand, aim to produce a "portrayal of a particular, more or less protracted occurrence of a unique, temporally circumscribed reality" (Windelband, 1998, p. 12). Windelband's *idiographic* describes an individual instance of something that emerges as the result of a particular methodological approach, or the individual that becomes a possible object of observation only under certain methodological circumstances. It is not a technical method, nor a kind of object in the world to be studied, but rather a research orientation that implies epistemological interdependency between observer, method and object (cf. Quine, 1951).

The distinction was publicized in the psychological domain by Hugo Münsterberg in his 1898 presidential address at the American Psychological Association meeting (Hurlburt & Knapp, 2006). The term gained further prominence in psychiatry as it was taken up by Gordon Allport in his seminal work on psychology of personality. Both Münsterberg and Allport suggest, following Windelband, that while the idiographic and the nomothetic are distinct as

[2] When the WPA sought out to develop guidelines for diagnosis in 2003, they established a working group for "Idiographic (Personalised) Diagnostic Formulation" (Thornton, 2007).

theoretical categories, they are also practically commensurable (Windelband, 1998, p. 19). All successful description, let alone psychological research, requires both kinds of investigation:[3]

> The separation of the single facts from the general facts is … untenable, because the explanatory law includes the description; but we can also emphasize the other side of this mutual relation: every description includes explanation, every assertion of a special fact demands reference to the general facts. (Münsterberg 1899 pp. 8)

Despite this acknowledgement by Münsterberg , the distinction became, by the mid-20th century, the basis of an opposition between two camps of practitioners (Salvatore & Valsiner, 2010).

Playing a role in this shift toward an oppositional understanding of the terms was the fact that nomothetic became synonymous with a more particular set of methods that were not indicated by the term when it first came into use at the turn of the 20th century. Allport popularized the term in the context of then novel methods in the nascent "trait approach" to the psychology of personality which, following the "Galtonian paradigm" developed by Francis Galton for intelligence testing and ultimately eugenics, involved controlled experiments with large sample sizes which were interpreted according to then novel statistical methods that are now common, including regression analysis and significance testing (eg. with ANOVA) (Robinson, 2011). These specific methods became known as *nomothetic*, even though Allport himself recognized that this term was not originally meant to refer to any particular method. Today the term *nomothetic* is commonly used to describe studies that use large sample sizes in the aggregate to demonstrate statistical significance (Salvatore & Valsiner, 2010). This notion of

[3] This view from Münsterberg, and its emphasis on description, reflects the kind of description Hegel venerates in *Phenomenology of Spirit:* "To judge a thing that has substance and solid worth is quite easy, to comprehend it is much harder, and to blend judgement and comprehension in a definitive description is the hardest thing of all" (Hegel, 2013 [1807], p. 3). This should perhaps come as no surprise given the overt Hegalianism of Windelband, the originator of the terms idiographic and nomothetic (Hartung, 2015) (for Hegel's account of the individual and universal in psychological research see Hegel 2013 [1807] pp. 178-185).

nomothetic is narrow compared to its original usage which referred to the inevitable role of generalization in even the most basic of scientific descriptions, which always occurs alongside the idiographic. The complex notion of the *nomothetic* approach which included a particular methodological orientation as well as a kind of object of study that emerges through that orientation was, over the course of the 20th century reduced to signify a set of methods. The breadth of the original notion of the *nomothetic* and *idiographic* is largely absent from contemporary discourse on psychiatry and machine learning.

At my field site, the term *idiographic* is used to describe the methodological choice of making use of only the participant's own data in the analysis on which recommendations for that participant are based. This follows a methodological definition of idiographic common in contemporary psychiatry: "personality traits and types allow us to compare one person with another…[in] the nomothetic approach…Alternatively, some psychologists study personality without focusing on individual difference. The idiographic approach studies individuals one at a time, without making comparisons with other people" (Cloninger, 1996 qtd. in Lamiell, 1998, p. 24). Proponents of the approach suggest that what may have once been seen as complicating eccentricities that make patients unique can now be harnessed for developing and enhancing "person-centered care" (Verhagen et al., 2022). The idea of *idiographic* and *nomothetic* as referring to sample size and interparticipant analysis is also present in the literature on machine learning. In a discussion of human and artificial intelligence and machine learning, Pedersen & Johansen describe nomothetic investigations as those that seek "insight into the distribution of preferences in a larger group or a larger population" (Pedersen & Johansen, 2020). They go on to suggest that this kind of study can answer questions about "tendencies on a group level," but might not make the correct prediction for a particular individual.

In these contemporary cases, what are sharply delineated as idiographic and nomothetic investigations are, in the way Windelband and Münsterberg knew these terms, mixes of idiographic and nomothetic. Both processes accumulate discrete data points and aggregate them for the sake of a generalization expressed as attributes of interest plotted against an independent variable. The size of the sample and degree of comparison across samples is what distinguishes that which, in the contemporary lexicon, is known as *idiographic* and *nomothetic*. As the term was originally explained, no contemporary psychological study nor psychiatric intervention could be wholly idiographic.

The term *idiographic* as its used in the TPL is also, in a way, unique from the mainstream contemporary meaning. In its current usage, the idiographic is regularly, sometimes overtly and sometimes by implication, connected to meaning, discourse and narrative. A group of scholars focused on "advancement of the idiographic science," including the editors of the *Yearbook of Idiographic Science*, Sergio Salvatero and Jaan Valsiner, have developed a research program involving integrating psychiatry and semiotics, considering sense-making the fundamental subject of idiographic psychology (Salvatore & Valsiner, 2008). In a demonstration of institutional backing of the idiographic approach, the WPA convened a working group on the topic in 2003. As is the subject of Tim Thornton's (2007) critical article, the working group contributed to a trend of conflating idiographic with narrative, emphasizing the "individual story" of each patient. Further, when the concept of the idiographic appears in the pages of *AI & Society* in a 2020 article on "Behavioural Artificial Intelligence," the coupling of narrative and idiographic knowledge reappears: "If we are interested in gaining knowledge about how one specific person describes the driving forces of his or her life – that person's life-narrative – we will of course gain valuable knowledge about how this person understands his or her life and

ascribes or interprets meaning into it. This type of investigative approaches – generally known to lead to idiographic knowledge – would provide valuable insight into a specific, local and unique phenomenon" (Pedersen & Johansen, 2020, p. 527). In these cases, the idiographic seems to be, even if incidentally, closely tied to the notion of narrative—at least enough to warrant an article explicating the conflation of these terms.

The trial at the TPL, however, describes itself as idiographic but makes no mention of narrative, symbolism or first-hand accounts of experience. In another domain that constitutes a key part of the lab, machine learning and mobile app development, personalization takes on a different conceptualization that is relevant to TPL's development of a personalized psychiatric intervention using machine learning and a mobile app.

2.2 Recommender systems, pesronalization and the user

While the TPL's trial is idiographic in the sense that it uses an N-of-1 sample size, it does not include the narrative or semiotic approach common to many contemporary investigations that describe themselves in this way. Following from this indifference toward personal narratives, the TPL trial does not include qualitative data and also does not forego generalization or the use of categories. The kind of personalization at work in the TPL study can thus be understood in the context of recommender systems, a field that includes computational systems that aim to predict user preferences.

Recommender systems are based in the effort to deliver personalized results to a wide range of users. A working notion of the individual is essential to designing these systems, as they are built to respond to the challenge of serving individuals with unique content to the extent that they are unique from one another. One example of a recommender system is music recommendation. The Pandora music streaming website and application takes an approach to the individual that closely reflects the TPL trial.

In a comparison of Spotify and Pandora recommender systems, Robert Prey describes Pandora's personalization project as based on the notion of an individual who is continuously changing over time, whose individual data would only be impoverished by integration into a population dataset. Pandora assesses individuals for the *musical attributes* commonly found in the songs they choose to listen to. This approach explicitly avoids analyzing the individual as a member of a group and does not use the choices of similar individuals to provide recommendations—tactics that are a key part of the recommendation strategy of Spotify, the industry leader in music streaming. Pandora's individual user is produced by continuously quantifying behaviors (mostly music streaming choices), and on the basis of that single user's behaviors alone, offering recommendations that are not sullied by the "cultural influence" of inter-individual comparisons or statistical modeling.

After deploying this highly individualized system, Pandora still must align with the workings of the advertising industry on which it depends for revenue. Advertising in music streaming is based on demographic categories. Only at this stage does Pandora use large group sampling to determine the kinds of musical choices that are common among a demographic. Robert Prey uses the example of Hispanics. To sell a "Hispanic" audience to advertisers, Pandora first identifies some users as Hispanic based on some non-musical information like a zip code. Pandora then quantifies the musical attributes typical found in this community's chosen songs. On that basis they can infer other people's Hispanic-ness, using music taste as a proxy. Thus, for users outside of the small sample group whose zip code was analyzed, Hispanic-ness, in the eyes of Pandora and their advertisers, is a matter of song choice; " 'Hispanic-ness' is performed into being through music listening behavior" (Prey, 2018, p. 1094).

Pandora's personalization system is similar to the TPL trial's idiographic personalization. The individual in the TPL trial becomes available to data scientists as a collection of behaviors. Reluctance to compare the individual's data to that of a larger collection of individuals is promoted as a feature of the system, which honors the unique nature of each individual who otherwise is likely to be "pigeon-holed" into a generic category. This highly personalized and dynamic prediction system produces, as its output, one of four categories: diet, exercise, positivity, or sleep. These categories are importantly "post-hoc," however. The ML system itself does not classify users into these broad categories, but rather produces a more rich analysis of to what degree each of the measured variables (eg. walk speed, positive mood, daily sugar ingested) relate to one another statistically. This aspect reflects the fact that, in the Pandora case, rather than make comparisons across users and learn similarities of user behavior as demographic categories in the first place, the system analyzes a user for *their choices of musical content over time*, producing an individual uncorrupted by the masses. This unique individual is then labeled as belonging to a demographic to the extent that it meets some standard of demographic behavior.

Similarities continue into the latest initiatives from companies like Pandora and Spotify which, in what has been called the "contextual turn" in recommender systems (Pagano et al., 2016), have taken increasing interest in the activities of daily living that constitute the "context" of digital music streaming. Pandora and Spotify both aim to modify their recommendations according to an individual's behaviors including not only song choices but also movements from place to place, and other activities that can be measured by smartphone sensors. While it initially appeared that Pandora and TPL are distinguished by the kind of data that they use and value, in fact Pandora aims to make use of similar data to that used by TPL in their trial. In another

similarity between music recommender systems and the TPL, the "contexts" that Spotify aims to detect in order to optimize recommendations overlap considerably with the foci of TPL's behavioral modification plans. Spotify's contexts are "chill time," "workout," "party," "dinner," "focus," and "sleep" (Prey, 2018). These contexts are seen as exhaustive for Spotify, which aims to provide recommendations for *all of the contexts* in which a user might find themself at any moment. Sleeping, eating and exercising are relevant activities as "contexts" in Spotify's system and as "plans" in the TPL trial. These "contexts," for Spotify, constitute a branding opportunity to serve particular ads in each situation: "Bacardi, Gatorade, and Bose were among the big brands making up the 'select launch partners' for 'Branded Moments': Bacardi on 'Party', Gatorade on 'Workout', and Bose on 'Chill' " (Prey, 2018, p. 1094).

While these similarities between recommender systems and the TPL trial could be described as a shared interest in "context," the idea of context varies considerably by the method used to study whatever primary object is contextualized, and by the theoretical orientation of the investigation itself. Therefore the question is not so much about a shared interest in context, but a strikingly similar notion of *what context is* (Seaver, 2015). Listening experiences are seen as influenced by contextual factors like sleep and exercise, as are experiences of depressed mood at the TPL. Both systems emphasize the power of considering these contextual factors in improving user experience.

2.3 Personalization and context in precision medicine

These ideas of personalization and context can be productively contrasted with the notion of personalization common in precision medicine (PM). This field aims to leverage large datasets (largely produced and operationalized in genomics, proteomics, metabolomics, or some other field ending in -omics) for each of a single patient's many "risk-factors" or "indications." A single patient's genetic predisposition to myriad diseases can be determined by analyzing their

genome within large datasets designed for each disease. This results in a unique individual profile that expresses risk to a range of genetic diseases which could then serve as a basis for personalized lifestyle or therapeutic interventions.[4]

The precision medicine approach is distinguished from the personalized approach of the TPL in that it iteratively locates the individual as one among many, using the individual's similarity or dissimilarity to others (who have particular health outcomes) as its primary means of individualization (Baumgartner, 2021). This is altogether unlike the TPL's approach, which refrains from comparing individuals at all, and only groups individuals together at the end of the investigation when a behavior plan constitutes its output[5]. The PM approach considers the individual as already among the collective, in need of individualizing by means of quantifying differences among members of the population. In this sense, PM's individuality is defined negatively—it becomes meaningful only by seeing the individual among the collective. The trial design at TPL, on the other hand, aims to constitute the individual from scratch, ignoring what could be taken as "comparable" individuals. Thus the question is raised: if not by its difference from the rest, how does this individual of the TPL become meaningful?

Along with a unique statistical notion of the individual comes a unique sense of "context." By offering multi-dimensional measurements of patients' genetics, the PM program can also offer context, which is characterized by the logic of statistical correlation across individuals and across indications (conditions, diseases or phenotypes). Individual risk can emerge from the cooccurrence of a few genetic factors that is commonly found to correlate with

[4] Determining a patient's risk for a high number of conditions results in such a unique profile that the diversity of patient profiles becomes unwieldy and even threatens patient confidentiality, as a highly tailored disease risk profile or treatment plan would be the result of only one particular combination of risk levels. This problem is called the curse of dimensionality (Berisha et al., 2021).

[5] Despite this distinction, former director of the National Institute of Mental Health in the U.S., Thomas Insel, describes digital phenotyping as "precision medicine for psychiatry" (2017, p. 1215).

disease in the greater population. The individual, in this case, is elaborated through manifold comparison to peers. A powerful insight from PM is one which makes the right comparison, showing how certain aspects of the participant also occur in another group that often develops some health outcome. This is distinct from the TPL's approach, which offers insight by showing a correlation in time between the event of a particular behavior and a health outcome. The TPL's individual is elaborated by selecting relevant events or moments in time which bring together behavior and mood outcome. The PM approach elaborates the individual by selecting relevant permanent attributes of the individual and comparing them to a group on the basis of these attributes.

Notions of personalization and the individual participant are at the center of the disciplinary history of digital psychiatry and digital phenotyping, as well as the contemporary landscape of related recommender systems. The TPL is idiographic in one contemporary sense of limiting its analyses to a single participant, but breaks from the trend of idiographic investigations soliciting personal narratives. It aligns with the kind of personalization at work in some recommender systems, which generate recommendations by analyzing a user's behaviors in isolation and generate group data, when necessary, post-hoc. These systems also group together the relevant activities of life as "context" categories that reflect the PWP plans at the TPL. This notion of the individual and its relation to narrative and context are in contrast to precision medicine, which gets its strength from presenting an individual in terms of their relative position in a large distribution of other individuals. As we have already seen to some extent, the individual studied at the TPL and the methods required to personalize the intervention for each individual is closely bound up with the kinds of data the laboratory uses and the methods it uses for collecting this data. Of particular relevance is the distinction between

meaningful, symbolic or narrative accounts of experience and more prosaic kinds of behavioral data.

3. Meaning and Meaninglessness: From Psychoanalysis to CBT to PWP Guidance

From the first announcement of the trial, the PI consistently assures lab members that guiding participants is *not therapy*, suggesting that a similar study could use chatbots instead of human guides which, while sufficient for the primary purposes of the study, would lack a "personal element." The PI even goes as far to say that participants do not need the guide to bring about change in their mental health because those changes are the result of the behavior—the real "intervention" of the study is the participants' change in their own behaviors, not the guidance. To understand this kind of intervention, which lab members describe as similar to CBT and yet "not really therapy," it is necessary to understand how the history of CBT is shaped by an opposition to psychoanalysis.

3.1 Psychoanalysis

Psychoanalytic psychotherapy is today unified, to some extent, by its focus on the unconscious as its primary therapeutic object, and on language and meaning as principle theoretical structures. Both of these characteristics of psychoanalysis extend back to its Freudian origins. Freud considered language as constitutive of thought, and described it as in opposition to action, prescribing prohibitions on actions and utterances and describing the clinical situation as "struggle between the doctor and the patient, between intellect and instinctual life, between understanding and seeking to act . . . " (Freud, 1999, p. 108 qtd. in Stern, 2002).

Both the importance of the unconscious as well as the guiding structure of language carried through the work of Lacan who, in the 1960's became a controversial and influential voice in the field, famously proclaiming that "the unconscious is structured like a language." The unconscious, itself the primary phenomenon under investigation in psychoanalysis, includes the

three other phenomena, each of which is a manifestation of the unconscious itself. These are repetition compulsion, transference and the drives, schematized as such in Lacan's *Four Foundational Principles of Psychoanalysis*. Following the structuralism of Claude Levi-Strauss, Lacan describes the unconscious as a pre-subjective network of signification into which the subject situates themself. Rather than produce the subject position in the act of uttering the "I" of a first-person statement, the subject must grapple with a pre-existing linguistic structure from the start. This "I" of grammatic predication is not the same as the ego of first-person experience. The unconscious presents itself as the problematic by which the I and the ego are separated, which may bring about a stutter or some other failure or reluctance to speak (Lacan, 1998, p. 20; Stern, 2002, p. 222). Lacan locates the unconscious in this "gap," "lack" or "discontinuity" which he sees in the difference not only between the "I" and the ego, but also the ego and its self-image, and the cause and its effect (1998).

Lacan's emphasis on language and meaning endures in today's psychoanalytic practice, which is considered by some practitioners an "intersubjective dynamic of co-constructed meaning." Salvatore et al term this a "dialogic and semiotic approach," and suggest that it is shared across varieties of psychoanalytically inflected therapies (2010, p. 196).

The epistemology of contemporary psychoanalysis relates to the contemporary notion of idiographic—it has come to be associated with individual and the way that individual "formulates" their life: "The object of psychoanalysis is not the human object is general, but the human object in so far as he himself formulates his own experience" (Laplanche, 1987, p. 13 qtd. in Lussier 1991). Preferring the qualitative and interpretive case study, Freud was reluctant to accept Saul Rosenzweig's attempt to conduct controlled experimental trials on repression, suggesting that results from such a method could only confirm, but could not disprove the

psychoanalytic notion of repression (Lussier, 1991, p. 61). Prominent psychoanalyst Erik Erikson explains the importance of focusing on individual cases in analysis by describing the patient as a "universe of one" (qtd. in Schön, 1983, p. 108).

While Lacan's work is well known as a theoretical point of departure for postmodern philosophy and hermeneutics (eg. Deleuze & Guattari, 1983, pp. 38–44; Ricœur, 1970) it also, in the spirit of Lacan's practice-oriented definition of psychoanalysis and his own commitment to the clinic, serves as a canonical guide to therapeutic practice (eg. Fink, 2007). The activities of practice that unify contemporary psychoanalysts include framing, presence and engagement. Framing consists of maintaining the roles of therapist and patient as well as the professional therapeutic environment. Presence involves the therapists' attunement to their patient as well as their mirroring of patient behaviors. Finally, engagement describes the verbal and nonverbal communication in which therapists and their patients partake. (Karbelnig, 2022).

While most psychotherapists employ these aspects of clinical practice and some attend, at least to some degree, to Lacan's Foundational Principles, his work is not only a source of unifying theory but also a contentious point of disagreement—the lectures that would become *Four Principles* immediately followed his excommunication from the International Psychoanalytical Association. Other influential psychoanalysts including Jung and Adler had similar fallings out with the institution in the 1960's (Lussier, 1991, p. 58). With each of these analysts going on to develop their own psychoanalytic programs, the "common ground" of the discipline has been in question since at latest the 1980's (Karbelnig, 2022; Schafer, 1985). In an effort to maintain and explicate the qualities of psychoanalysis that make the discipline whole and unified, André Lussier proposes resistance to a scientific program that identities itself "with the 'exact' and the 'measurable' instead of with the 'meaningful' " (1991, p. 60). The scientific

pursuit of the exact and measurable that is rejected here as a threat to the unique disciplinary identity of psychotherapy is welcomed by cognitive behavioral therapy, the increasingly dominant therapeutic camp that is often presented as the evidence-based alternative to psychotherapy.

3.2 Cognitive Behavioral Therapy

The behavioral modification plans that constitute a significant part of the intervention at my field site are based on studies that, in both explicit and implicit ways build on the approach known as cognitive behavioral therapy (CBT). The above historical and epistemological sketch of psychoanalysis is necessary to understand CBT. The approach emerged out of a crisis in psychoanalysis, at which point the two camps defined themselves in opposition to one another in what Norcross and Newman called "a dogma eat dogma environment" (1992). While today CBT is often venerated for how it exceeds psychoanalysis in terms of scientific rigor[6] (see Shedler, 2011 for a critical account), a look at the development of CBT shows the close conceptual relationship between the two approaches.

While the early twentieth century saw the rise of modern behaviorism, popularized in the work of Watson (1970 [1924]), and its development in the early years of psychology by neobehaviorists such as Skinner (1965), these theories made little impact on psychotherapy which was, through the 1960's, dominated by psychoanalytic and psychodynamic approaches. As these psychiatric camps were assimilated into debates on structural problems with the social environment, they came to be known under the banner of the psychosocial model (Wilson, 1993).

[6] The comparative clinical efficacy of these therapies is beyond the scope of the present dissertation.

The eventual dissolution of this approach and the decline of psychoanalysis that came along with it has been attributed to a wide range of institutional, political and scientific factors. Among these are the growth of the medical paradigm or "biomedical model" for classifying illnesses, which proliferated through the trailing of clinical psychologists under the "Boulder model." The decline of psychoanalysis and the proliferation of the biomedical model coincided with not only the reduction of the term *nomothetic* to reference to the large personality-trait studies like those of Gordon Allport, but also a formulation for medical training based on the two distinct components of research and practice, the former a superior guiding structure that is merely implemented in the latter. The field of psychoanalysis, according to Donald Schön, exemplifies an exception to the technical rationality that divides medicine into generation of theory on one hand and the "application" of this theory in practice on the other. Perhaps unsurprisingly given its resistance to this positivistic division of labor, psychoanalysis was deemed unscientific by many psychologists in field and even called a pseudo-science by Karl Popper (2002). The alternative biomedical model gained prominence due to the criticism that psychoanalytic and psychosocial diagnosis fails to adequately distinguish the well from the sick—a criticism which, by the publication of its third edition, could not be leveled against the Diagnostic and Statistical Manual of Mental Disorders (DSM-III), a key instrument for biomedical psychiatry (Wilson, 1993). The biomedical model also sought diagnostic specificity to rebuke the anti-psychiatry movement of the 1960's which questioned the existence of neurotic and psychotic conditions (Paris, 2005, pp. 83–85). Antipsychiatry ideology among both right and left leaning political groups related to an anti-authoritarian ethos that has been attributed as a broad societal factor in the dethroning of psychoanalysis as well as a personal motivation for the

founders of two of the key therapeutic forerunners of CBT, Aaron Beck's cognitive therapy and Albert Ellis' rational emotional therapy (Rosner, 2014; Wilson, 1993, p. 402).

Ellis outlined his rational-emotional therapy (RET) in *Reason and Emotion in Psychotherapy*, which was first published in 1962. Ellis, though a former analyst, explicitly opposed psychoanalysis. He developed RET because of his feeling that psychoanalysis, even in cases in which it yields insight that patients acknowledge as plausible and profound, fails to deliver the desired result of "minimal anxiety and hostility."

Ellis considers his development of RET a result of the essential agreement between Freudianism and Pavlov's theory of conditioning. He suggests that what Freud calls the Oedipal fear of the father is an instance in which the patient is "conditioned early in his life to fear something," like Pavlov's dogs, who were famously conditioned to the sound of a dinner bell such that they salivate at the mere sound of the bell (Ellis, 1963, p. 10). He goes on to suggest that, while it is reasonable to suspect that a human's conscious knowledge of the fact of this conditioning would break it, his psychoanalysis patients show little progress despite their recognitions of relevant past trauma. Ellis looks back to Pavlov, whose deconditioning process involved ringing the bell in the absence of feeding which eventually broke the conditioning. Ellis' application of this idea to his human patients resembles what is known today as exposure therapy: expose the patient to the object that inspires fear of a bad outcome; when the bad outcome never comes, the patient will no longer be conditioned to relate the object to the bad outcome and will be free of fear. Ellis claims that, recognizing the limitations of this strictly behaviorist model, he then realized the fundamental difference between dogs and his humans subjects:

> Human beings have one attribute which none of the other living beings that we know have in any well-developed form: language and the symbol-producing facility that goes with

> language (Cassirer, 1953; Whorf, 1956). They are able to communicate with others and (perhaps more importantly, as far as neurosis and psychosis are concerned) with themselves in a manner that is infinitely more complex and variegated than is the signaling of other animals.[7] (Ellis, 1963, p. 14)

Ellis goes on to explain that humans can be conditioned not only by sensations, which are similarly felt by dogs, but also by the *definitions* given to things through language. His ideas on language are strikingly simple compared to those that Lacan was developing in the same historical moment. His interest was in "the precise, simple declarative and exclamatory sentences" that defined objects in affectively charged terms. The transcripts used as examples in Ellis' writings have been "clarified," as their verbatim contents were, in Ellis' view, "unwieldy, discursive and… somewhat unclear." As the analyses of the transcripts demonstrate, Ellis takes little interest in the "discourse" of his patients' speech, but rather reduces it to a simple definitional or syllogistic form. For example he suggests "I am worthless" as the definitional simplification of the self-reflections of a patient suffering from depressed feelings. He then disputes this proposition as illogical (on the basis of its tautology, for instance) or indefensible (as the result of, for instance, its evidence having been subjected to selection bias) (Ellis, 1963, pp. 29–33).

To recap, Ellis begins his theory with a perplexing suggestion of equivalence between Pavlovian behaviorism and Freudianism; he then moves to add to Pavlov's theory a consideration of language. He goes on to supplement Pavlov's theory with the notion of a critical

[7] References here are to Ernst Cassirer's *Essay on Man* and Benjamin Lee Whorf's *Language Thought and Reality*. The latter contains papers that became known for the controversial hypothesis of linguistic relativity. Stephen Levinson's forward to the 2012 edition notes Whorf's relative dissatisfaction with psychological behaviorism as an example of how his thought was distinct from popular intellectual currents of his time (Levinson, 2012, p. ix). Ellis' behaviorism is here expounded by reference to a theorist of language whose eclecticism is exemplified by his rejection of behaviorism itself.

self-consciousness that sounds like Freud's notion of ego and superego transposed into the domain of conscious thought: "as man's complex or secondary self-signaling processes arise, a new factor comes into play that may enormously change the simple going-toward or avoidance equations made by lower animals. This factor may be called self-consciousness or thinking about thinking" (Ellis, 1963, p. 15). According to his own account of how RET developed, Ellis abandoned psychoanalysis for the, at the time diametrically opposed, camp of behaviorism. He then reintroduced to his behavioral model notions of the of language and of self-consciousness, as if having discovered them anew.[8] While these might have been similar to Freudian ideas, in Ellis' new theory they were no longer held in the realm of the unconscious, but rather that of conscious thought, and the way they changed over time and responded to treatment followed not from the dynamics of the unconscious structures of desire such as drives, but rather through the rules of logic and empirical verification.

By the time Ellis had published his fully developed theory of RET, he had become friends with Aaron Beck, the psychiatrist now largely credited with the CBT approach. Beck, like Ellis, was trained as a psychoanalyst and spent much of his career analyzing patients and conducting research in the Freudian paradigm. By the late 1950's, though, Beck had developed hypotheses that were contrary to the psychoanalytic dogma and was eager to test them. In his challenge to the notion, long held in psychoanalysis, that depression was "a wish to turn anger inward on one's self," Beck studied his depressed patients for the presence of hostility in dreams, the expressive domain of Freud's unconscious. Rather than abandon Freud and start fresh with Pavlov, Beck wanted to contribute to the growth of psychoanalysis.

[8] He writes: "I began to put all my psychological and philosophical knowledge together in a somewhat different way than I had previously done" (1963, p. 14).

Ultimately unsatisfied with the rigor of dream evaluations, he developed the still widely used Beck Depression Inventory—a diagnostic instrument that measured depressive thoughts rather than dream content (Clark & Beck, 1999, p. 46; Rosner, 2014). As an alternative to a Freudian etiology, Beck explained his results by describing depression as a "thought disorder." The potential threat to psychoanalysis posed by this perspective became less relevant as Beck's department at the University of Pennsylvania underwent a radical move away from psychoanalytic thought and its practitioners. In this sense, Beck's new ideas didn't attack psychoanalysis, but rather filled the void that emerged as psychoanalysis was successfully discredited by a group of otherwise intellectually disparate researchers and clinicians (Rosner, 2014).

As Beck developed his "thought disorder" ideas in 1963 and 1964 publications, he started to demonstrate the clear influence of cognitive psychology, characterizing depression as "a systematic error; viz, a bias against [the self]," going on to suggest that "the typical depressive cognitions can be categorized according to the ways in which they deviate from logical or realistic thinking" (A. T. Beck, 1963, p. 328). Unlike Ellis, Beck did not claim a novel theoretical basis for his ideas about thought disorders. Instead he attributed them to the existing psychological literature and suggested that they only seemed new to psychiatrists because of "a notable lag in the application of the structural concepts generated by studies of normal thinking to the thinking disorder associated with various psychiatric syndromes" (A. T. Beck, 1964, p. 562). From this psychological literature emerged a theory of concepts that would serve Beck's ideas about depression.

Beck's notion of the concept shared with Ellis' theory the emphasis on a logical and defensible syllogism, but was described in terms from the emerging field of cognitive

psychology. Ellis describes the logical component of his theory as "philosophical causation," referring to the philosophy of logic. Beck, on the other hand, explains the role of the syllogism as a kind of schema, citing the emerging literature on schema and concept in cognitive psychology. Specifically, he describes faulty schematic thinking through examples of deductive reasoning on the basis of a false syllogistic premise. The patient has the schema "all dogs will bite me." On this basis the patient deduces that a particular dog, which happens to be very tame, will bite them. The faulty deduction points to a schema or premise that is itself untrue. The deductive process, however, is perfectly logical, even in the pathological patient. The pathogenic, or in Becks' term, paralogical schema may be inactive during some nondepressed, asymptomatic periods. At these times, other schemas dominate. During depressive periods the paralogical schema becomes hyperactive, exceeding other schemas in "strength," despite potentially low "truth value" (A. T. Beck, 1964, p. 565). By explaining the syllogism in terms of schema rather than "philosophical causation," Beck's theory was made intelligible to the then emerging field of cognitive science.[9] His theory also notably distinguished the content of a thought from thought processes, suggesting, in opposition to behaviorist epistemology, that problematic or "paralogical" thoughts can lurk within the mind without being activated. CBT and its forerunners broke from behaviorism not by positing a theory of meaning, but rather an underlying cognitive aspect of the self that may not always be manifest. In this way, as is discussed in Chapter 3, Beck lays the theoretical groundwork for the way lab members conceptualize the trial at the TPL.

[9] In the 1964 paper that first explained the theory of CT, Beck cites *A Study of Thinking* (J. Bruner et al., 1956) as an example of what he means by "schema." Eleanor Rosch would go on to cite Bruner et al as an example of "the classical view" of concepts in cognitive science (Rosch, 2011, p. 92). By the time of his (1990) publication *Acts of Meaning,* Bruner had become a critic of the computational metaphor for cognition and a proponent of considerations of narrative in thought.

Beck originally aimed for his novel approach to offer common ground for psychoanalysts, clinical psychologists and behavioral scientists. Despite the now common framing of CBT as the evidence-based, scientifically verified alternative to psychoanalysis, Beck saw his CT as largely commensurable with psychoanalysis (Rosner, 2014). The two fields share conceptual terrain. Both emphasize the linguistic: Beck considers language content a necessary feature of cognitions. Both CT and psychoanalysis use a linguistic model to theorize a structure of thought beyond empirically manifest symptoms. In later texts on the topic, Beck's schemas are called "meaning-assignment structures" or "meaning-making structures"(Clark & Beck, 1999, p. 65). On the topic of depression specifically, Beck's CT also shares with psychoanalysis some key diagnostic components and, unlike Ellis, Beck acknowledges this common ground. The negative paralogical cognitions regarding the self—a thematic pattern that distinguishes the false thoughts that cause depressive cognitions—is analogous to Freud's concept of self-esteem (Clark & Beck, 1999). This is just one example of how Beck, like Ellis, attempts to transpose Freudian ideas about unconscious desire onto the realm of conscious thought or the "cognitive."

While Ellis abandoned psychoanalysis and started from the basis of Pavlov's theory, which had by then become foundational to behaviorism, Beck more gradually broke with psychoanalysis and his work remained, for years after publishing his outline of CT, seen as incompatible with behaviorism. Beck's ideas were rejected by the behaviorist camp when CT was seen as a bridge from psychoanalysis. Only as part of the burgeoning field of cognitive science did Beck finally find support from behaviorist thinkers who were eager to jump onto the "cognitive science bandwagon" (Rosner, 2014, p. 756).

Despite Ellis' assertion that the theories of Freud and Pavlov were not so different in their essence, neither CT nor RET were ever reconciled with the stricter form of behaviorism common

in mid 20th century psychology. Behaviorist thinkers did develop theories of and therapies for depression, such as those based on the stimulus-response model, which foregoes intermediating thought structures in favor of a model of the mind based only on observables inputs and outputs. When these behaviorist psychologists turned to depression they defined it as a low rate of response-contingent positive reinforcement. The thoughts of low self-esteem and other negative cognitions that are central to Ellis' and Beck's new therapies were, in this camp, seen as downstream epiphenomena that were largely irrelevant to the treatment program (Friedman & Katz, 1974).

Behaviorism in psychology and psychiatry was, by the 1970's, on its way out as the cognitive revolution got underway. Today's cognitive behavioral therapy continues along the lines of Beck's 1960's publications, consisting of an approach to psychiatric treatment that emphasizes cognitions and cognitive structures as the causal root of affective issues (Clark & Beck, 1999, p. 37). Its treatment consists of brief courses of therapy in which the therapist assists the patient in intentionally modifying these patterns of thought, on the basis that changes in thinking lead to changes in behavior. Patients may also be directed to modify behaviors directly. Other key attributes of contemporary CBT include a sound, collaborative, educational relationship between therapist and client; a structured and time-sensitive plan toward achieving specific goals; highly structured therapy sessions; the use of worksheets and "Action Plans" as homework; and a reluctance to let focus shift to past events. Within this structure suggestions for cognitive-behavioral modifications can be wide-ranging, and even taken from other kinds of therapy such as psychodynamic therapy or Gestalt therapy (J. S. Beck, 2021).

Some forms of CBT are based mostly on modifying action and behavior, rather than patterns of thought. Three subfields of CBT, coping skills therapy, problem-solving therapy and

cognitive restructuring therapy are distinguished largely by their different mixes of cognitive and behavioral modification programs. However, for an approach to qualify as a CBT, it must demonstrate a cognitive "mediation" between behavior and outcome. Strict stimulus response or operant conditioning approaches therefore do not qualify as CBT (Dobson & Dozois, 2010). Because the empirical verification of this mediation is not a well-established part of the CBT literature,[10] which nonetheless holds as a theoretical tenant that *behavior can only change as a result of cognitive change*, the mediational requirement consists largely in a theoretical orientation or a heuristic for explaining how observed behaviors relate. Therefore, a CBT approach can offer treatment suggestions that are entirely based on behavior and also measure outcomes entirely on the basis of behavior, while still qualifying as CBT because of a supposed underlying mediational mechanism at play.

While disagreements between psychoanalysis and CBT largely center on theories and practices of talk therapy, the late 20th century also featured a decline in the importance of talk therapy within the psychiatric discipline. Today, psychiatric science and therapy is largely concerned with pharmaceutical interventions and talk therapy is more often conducted by mental health professionals from other disciplines (Harris, 2011).

3.3 PWP guidance at The TPL Trial: Like CBT But Not Therapy

As it was originally planned and promoted, the trial at TPL includes a coaching or guidance process which, because lab members consider individual behaviors the crux of their intervention, is seen as a minimal and supplementary component. The talk that goes into explaining how and

[10] Dobson and Dozois suggest "tests of cognitive mediation are often less than methodologically adequate, and many fail to produce compelling results" (2010, p. 5). While put here as an incidental inadequacy resulting from a lack of rigorous investigation, Wittgenstein (1968) raises problems with inferring cognitive mediators that underlie terms like "expectation" (an example used by both Wittgenstein and Dobson & Dozois) which, he demonstrates, only gain sense from the broader pragmatic and grammatical context.

why the patient should, in the sleep plan for example, go to sleep at a different time, is not an essential part of the intervention. For this reason, the guidance is not considered therapy. While the PI suggests that it is "like CBT," and takes its sleeping plan recommendations from the subfield of CBT that deals with sleep, the lab is, overall mostly oriented toward modifying behavior and makes no mention of thoughts as an object of inquiry or intervention. As for the thoughts or cognitions at the heart of some forms of CBT, these become relevant in the solicitation of questions about mood in the EMAs as well as the positivity plan intervention, but are not explained as having a causal or mediating role in the intervention.

The TPL trial rejects therapy itself in a discourse that is similar to CBT's contemporary rejection of psychoanalysis, but goes even further. The TPL study makes use of the behaviorist theoretical underpinnings of CBT while rejecting what unites CBT with its psychoanalytic predecessor. Much of the cognitive theory behind CBT, as is clear in Ellis' and Beck's work, reinterprets concepts common in psychoanalysis in a new cognitive light. In this re-envisioning, patterns of conscious thought are the domain of cognition, not the unconscious. Both CBT and psychoanalysis take language and discourse as essential components of psychical processes. The supposed founders of both fields, Beck and Freud, consider thoughts as constituted linguistically. The TPL trial, on the other hand, takes no interest in the symbolic or logical relations between the behaviors, moods and other components of its study. The TPL does, however reference CBT as a related approach[11].

[11] Former NIMH director Thomas Insel further establishes the link between digital phenotyping and CBT's replacement of psychoanalysis by citing, in his article promoting digital phenotyping, Leon Isenberg's 1986 Eli Lilly Lecture to the Royal College of Psychiatrists, which laments a dearth of consideration of "social meaning," as, following the downfall of psychoanalysis, psychiatry has gone from "brainless" to "mindless" (Eisenberg, 1986). As we will see in Chapter 3, digital phenotyping at the TPL does not reintroduce the complexities of multiple meanings, but rather, in a related move, multiplies versions, layers or "ontologies" of the individual.

The similarity between the TPL trial and CBT is most clear in the case of Aaron Beck's initial attempt at devising a method to implement his ideas about the psychiatric risk of "paralogical," "negative" or "automatic" thoughts. Beck indulges in the self-reflexive practice of devising his methodology on the example of his own psyche:

> Twice a day, often more, [Beck] sat down with pen and paper and identified, classified, quantified in percentiles the degree of his belief in, and restructured every possible "negative" thought (he now calls them "automatic thoughts") he could capture on paper… These exercises shored him up to face each day with a sense of mastery over his emotions, immersed him in the real consequences of using his new cognitive approach, and helped him flesh out his new theory. … It is difficult to describe in words the explosion after 1962 of spiral-bound notebooks, inside front and back covers of notebooks, pocket notebooks, steno pads, 3-by-5 notecards, and random pages on which line after line of his "negative" thoughts are jotted down systematically, quantified in percentiles, classified (he toyed with different color-coded schemes), and reconfigured. (Rosner, 2014, p. 755)

The scheme that Beck developed through this self-analysis became the foundation for the CBT that is popular today. This kind of process is sometimes schematized through worksheets or homework assigned to patients by CBT practitioners (J. S. Beck, 2021, pp. 135–138). It closely reflects the digital phenotyping approach of the TPL. Lacking wearable or mobile phone technology, Beck used pen and paper to log his thoughts, attempting to bring attention to "automatic" patterns of thought that might otherwise go unnoticed. Beck's CT and the CBT that emerged out of it are distinct from the TPL trial, most importantly in their emphasis on the linguistic relations between thoughts and their inclusion of patterns of thought alongside patterns of behavior as therapeutic targets. That said, logging thoughts as events that occur over particular intervals of time is the exact method taken up by the TPL. It is also the practice that Aaron Beck, who is credited with founding CBT, made use of in the years immediately *preceding* the publication of his first writings on the topic.

This kind of self-tracking procedure is also a common practice among users who are not part of a formal medical intervention, but rather part of the quantified self movement. This movement consists in users voluntarily tracking their own behaviors and biomarkers, often using wearable devices like smart watches. Users justify their engagement in self-tracking practices citing a variety of goals: to objectively understand their own bodies and behaviors in order to make positive change, to gain new ways of seeing and relating to the self as an aesthetic and semiotic entity, to gain mindfulness, to resist social norms and even to provide support to their own narratives (Pols et al., 2019; Sharon & Zandbergen, 2017).

A technologically determinist interpretation of the TPL trial is foreclosed by this range of purposes for which users have taken up self-tracking, and by the fact that Beck used pen and paper for a process very similar to digital phenotyping. Other kinds of self-reflection projects can be undertaken using wearable sensors and other technologies can be used for the kind of daily logging of thoughts and behaviors common to Beck's CT and the TPL trial. That said, the quantitative basis of the movement has a profound effect on the ways it is practiced.

For QS users, the senses are inadequate instruments for keeping track of themselves. In their ethnographic inquiry into the movement, Pols and colleagues contrast QS from psychoanalysis, suggesting that "there is not a meaningful self" to be interpreted in the process of self-tracking. Instead, users "battle with a more prosaic unconsciousness of bodily facts, habits, automatisms and metabolisms of which the subject is not aware" (Pols et al., 2019, p. 105). This "prosaic unconsciousness" emerges from a protocol similar to that of Beck's CT, but is limited to behaviors and biomarkers and lacks a theory of thoughts as having relevant linguistic content. In Pols et al's investigation into self-tracking, they use the example of psychoanalysis as a point of contrast on the issue of meaning, furthering the notion that digital psychiatry is like CBT in its

opposition to psychoanalysis, and perhaps goes even further than CBT in this opposition, disregarding meaningful content of thoughts altogether. Looking at the QS movement and Beck's methodology side by side, we see the confluence of the contemporary trends around wearable devices, machine learning technology and the modern historical lineage of psychiatry.

The TPL trial thus appears as the latest movement in a trajectory that extends back to the decline of psychoanalysis in the late 20th-century. This trajectory is characterized by a progressive devaluation of interpretation and an increasing reluctance to posit the meaning of thoughts or psychological conditions. The TPL trial not only exemplifies this latest movement but also makes use of machine learning systems, the equity problems with which, we are told, are best solved through improved *explanation*. As disciplines collide, the question of meaning emerges as a central irony of machine learning approaches to the personalized behavioral intervention for mental health.

Rather than suggest a return to the interpretive approaches common in psychoanalytic and other talk therapy practices that were prominent in the mid-twentieth century, the present dissertation will demonstrate how attention to human interaction and interpretation invite considerations around the unpredictable kinds of contexts, including environmental factors, that emerge as people implement digital mental health interventions in the real world.

3.4 The Challenge of Explanation for Machine Learning and Digital Psychiatry

Issues of bias, explainability and accountability are common in social science and humanistic inquiries into machine learning because of the complexity of machine learning algorithms, as well as their "black box" tendency to produce inexplicable predictions after having been trained on labeled training data, in the case of supervised machine learning, or by unlabeled corpora, in the case of unsupervised learning. These questions can be epistemological, raising issues of how algorithms can be said to "learn" at all, and how they relate to human knowledge in cases in

which their outputs are inexplicable. These questions also have a normative dimension, in which relevant issues include the reproduction of racist or sexist biases, or difficulty in holding "black box" machine learning systems legally and socially accountable.

The insistence on improved explanations in machine learning is common in the realm of digital health, following the problem of distrust of AI in mental healthcare which supposedly results from a lack of "understanding of the underlying mechanism and how and when it should be applied" (Chandler et al., 2019, p. 11). Clinicians, for example, should have access to an explanation of ML systems that offers them an "output that can be tried to clinical constructs (eg. thought disorder)" (Chandler et al., 2019, p. 12). Here the "thought disorder" theory that distinguished Beck's early work on CT is called upon to serve as a post-hoc explanation for digital psychiatric interventions which have, in the case of the TPL lab and most other similar projects, given up on talk of "thought" altogether. Having explored the historical trajectory of psychiatry theory, which has lead the the TPL away from questions of meaning and interpretation and toward a more prosaic data-driven account, the challenge of explaining how ML systems work in this context no longer appears as an incidental oversight or technical malfunction, but rather as stemming from a fundamental incongruence in approach.

In the case of non-medical self-tracking, the trends, visualizations and other outputs from digital self-tracking app also need to be explained in order to take on permanent significance for users (Ruckenstein, 2014a). For several subjects studied by Pols et al in their self-tracking study, the interpretation and explanation of data traces constituted a fundamental component of the process of understanding the self.

4. Reflexivity in the Cognitive Sciences and Quantified Self

It is common for psychiatric interventions to consist of two processes: first learning about the particular mind of a patient, and then treating it. The convergence of epistemology and

therapy in the notion that self-knowledge itself constitutes a kind of treatment is a controversial current in psychoanalysis. Albert Ellis, mentor to Aaron Beck and founder of RET, found that psychoanalysis could help his patients make sense of themselves and find meaning in their relationships, but claimed this produced little therapeutic effect—an effect he expected to be built-in to the psychoanalytic method. The idea that knowing the mind is itself a treatment is, in a sense, the opposite to psychiatric interventions that suggest that some interventions may succeed despite the impossibility of their explanation (see, eg. Chandler et al 2019 pp. 11-12). The TPL's trial falls into the latter category and is therefore opposed to psychoanalysis in another sense: understanding psychoanalytic explanation can itself constitute therapy, whereas the TPL's therapy needn't include an explanation.

This notion of self-understanding as therapy is closely related to the question of reflexivity in the cognitive sciences: if epistemological achievements like learning about ourselves can constitute therapy, then it is natural to offer patients insights from epistemology of psychology itself, a field which includes understanding the psyche as one of its primary goals. Reflexivity here refers to the acknowledgement and study of connections between a scientist or clinician's own scientific methods and approaches, frameworks, cognitions and practices and those of a patient or participant they study.

The usurpation of behavioralism by cognitive science was aided by a reflexive argument that conflated good scientific thinking with descriptions of how the mind works (see above Section 3.2). The mind of the healthy person functions much like the cognitively-oriented psychologist: it constructs schemas, symbols and concepts and moves not just in response to stimulus but also in emergent flashes of insightful thought (Cohen-Cole, 2005). The reflexive move to merge scientific methodology with accurate description of a normal mind finds a

parallel in CT and the decline of psychoanalysis as well. The healthy CT patient is one who avoids the investigative mistakes that psychoanalysis was, at the time, often accused of making: logical fallacies, selection and confirmation bias, and simply coming to firm conclusions with a dearth of evidence. The healthy patient goes about everyday life as the cognitive therapist goes about their science, avoiding unhealthy logical leaps and statistical biases. In his first outline of CT, Beck describes the healthy patient as one who builds schemas in order to cope with social situations. In the healthy case, "the cognition resulting from the interaction of the schema with the stimuli may be expected to be a reasonably accurate (veridical) representation of reality" (A. T. Beck, 1964, p. 565). The healthy patient is not one whose responses to stimuli are adaptive to the environment, but rather one who accurately builds representations in the mind. Thus the reflexive imposition of the cognitivist way of doing science onto descriptions of everyday thought takes a step from the realm of psychology into that of psychiatry and psychopathology.

Further, the kind of cognition theorized by behaviorists like Skinner (1965), which consists in conditioned responses to stimuli, is not only a weak psychological theory but also a pathology. What Beck first termed the "negative" or "paralogical" thoughts eventually came to be known as "automatic thoughts"(Rosner, 2014). By intentionally controlling our thoughts to build logic schemas, Beck thought we could avoid mental illness. Skinner's "S-R model" was not just an inaccurate description of the healthy patient's mind, but a description of mental illness.

As he elaborates a description of the pathological patient, Beck's theory breaks from cognitive psychology of the time, instead opting for the explanations from logical positivism. The pathological patient does not suffer from underdetermined representations of reality or lacking symbolic or metaphorical systems of thought, but rather *logical* failures of *propositional*

thought that arise automatically. While Beck cites Bruner et al as he defines his schemas, a classic work in the early days of cognitive science, the schemas of CT are then explicated as propositions which can be fallacious according to, for example, the logical rules of the syllogism. The cognitivism-behaviorism debate saw participation from philosophers of science on both sides and logical positivism was closely bound to the behaviorist side (Cohen-Cole, 2005, p. 118).[12] Nonetheless, the relationship between methodology of psychology and description of healthy everyday thinking continues to be reflexive: the two merge into one another to the extent that Beck valued logic in his own methodology. While Beck's cognitive therapy was still primarily oriented toward ameliorating depressive symptoms, its effort to integrate itself with studies of the healthy mind is a precursor to translational psychiatry which explicitly combines research about minds and therapies on them[13].

As for reflexivity in the TPL trial, the fundamental epistemological orientation toward data-driven science is not prescribed as a healthy pattern of thought, in part because no theory of thought is involved in the intervention. Participants are known by their behaviors, not their thoughts, and therefore few parallels link them, in either healthy or pathological states, to the researchers who study them. However, reflexivity in another sense is a characteristic of the

[12] Beck's capitulation to logical positivism which, at the time was associated with behaviorism, is not so surprising given that his novel therapy was, despite the influence of cognitive science, a replacement *of psychoanalysis*, not of behaviorism. This alternative to psychoanalysis, as is made clear above, could not base itself on the importance of symbolism, metaphor and creativity, as these were central aspects of psychoanalysis (aspects for which the field was critiqued as anti- or non-scientific). That said, the psychoanalytic camp was not entirely averse to logical positivism (eg. Lacan's (1998, p. 198) theory of metaphor responds to the paradoxes of Bertrand Russell's (1920) theory of types).

[13] The emergence of the notion of translational medicine, with its distinct but parallel efforts to study the mind and simultaneously treat it, follows from the scientist-practitioner model (also known as the Boulder Model) in psychology, which set separate requirements for research and practice in Ph.D. training (referenced above in relation to the decline of the psychosocial model), as well as, more distantly, the 1910 Flexner Report which separated medical school curricula into basic science research and clinical training and subordinated the latter to the former (Albee, 2000; Schön, 1983, p. 28).

TPL's approach. Rather than a similarity between the theorized "healthy psychiatric patient" and "rigorous psychiatry researcher," in both the QS movement and the TPL's trial these roles collapse into one another. Patients themselves are important clinicians in the intervention. The digital app or the pen and paper constitutes the medium for self-reflexive behavioral modification, and if they are motivated and prudent in their self-tracking, they will see favorable outcomes. Rigorous scientific methodology and healthy behavior merge as patients are told to become prudent self-trackers as the first step in improving their health. One consequence of this role of the patient is the importance of living what is sometimes referred to as an "experimental life" (Kristensen & Ruckenstein, 2018). Because correlations between mood and behavior will be found using only the patients' own data, the more variation in behaviors they perform, the more opportunities to learn about changes in mood. For example, if a participant always sleeps only two hours each night, then changes in sleep duration will not be found to correlate with changes in mood, because the sample will lack any data representing longer sleep durations. To gain insight using this technology, it is important to develop a laboratory of personal behavior, testing out a range of lifestyles and behaviors.

This notion of self-reflexivity may seem to downplay the app and wearable device that help users achieve a quantitative understanding of their mood and behaviors. This raises questions about what roles the app and the user play in the TPL's intervention, and how to balance the agency of the user and the technology they use. The possible answers to this question differ according to how the therapy is characterized. In chapters 2 and 3 these questions will be explored in greater depth following empirical material from the TPL.

5. How does the present observational approach relate to CBT and psychoanalysis?

Just as the methods and theories of cognitive behavioral therapists and cognitive psychologists influence descriptions of their objects of study, ethnographers and social scientists also bring

perspectives that influence their engagement with and portrayal of their field site. The present observational approach sees psychiatric theories and the conceptual histories through which they become meaningful relevant insofar as they influence the practical activities and organization of the study. The approach also follows Ian Hacking's theory that concepts and categories can be constructed, but only in a world of activity whereby they are operationalized in ways that inevitably exceed formal accounts. It also draws on the field laboratory studies, a central contention of which is that the production of scientific knowledge is a situated practice (Lynch, 1993). This practice sees formal structures not as constraints on action but rather resources which scientists implement in myriad improvised ways in their day to day coping with data, technologies and one another.

In a monograph responding to a crisis of technical rationality in professional fields like law and medicine, Donald Schön (1983) develops a theory around the tacit, improvisational knowledge-in-action and following reflection-in-action that characterizes some areas of professional practice. Rather than an application of theoretical knowledge, this kind of knowledge escapes the theory/application dichotomy, coming into play most recognizably when formal concepts and categories fail to meet the practical needs of the situation (see also "torque" in Bowker & Star, 2000; Schön, 1983, pp. 49–69). Schön's theory built on foundational studies in interaction and phenomenology by Alfred Schutz and would go on influence the work of theorists of situated practice including Etienne Wenger (1990).

Schön takes the psychoanalytic training situation as an example of reflection-in-action. He demonstrates how analysts must forego prior schemata or theoretical frameworks in order to allow themselves to be surprised by what they hear from their patient or colleague. He shows how, through the concept of transference, the analyst sees his patients' emotional and

communicative issues as manifesting themselves within the emotional and communicative situation of therapy. When the analyst consults with a mentor or supervisor, the process of transference continues, and the practitioners consider how their own desires and the dynamics of their relationship relate to the problems of the patient and the therapeutic program. Through this concept of transference analysts are also encouraged to reflect on the relationship between the object of inquiry and the communicative situation in which it was brought up (1983, pp. 105–127).

The way analysts emphasize inter-subjectivity and reflection while refraining from applying theoretical categories, so as to embrace the improvisational character of the situation, not only serves as an exemple for Schön's knowledge-in-action and reflection-in-action, but it also reflects key attributes of the present approach, informed by ethnomethodology and laboratory studies. The similarities between psychoanalytic practice and the present observational approach justify attention to psychoanalysis as a foil which CBT and digital psychiatry position themselves against.

6. Conclusions and connections to later chapters

The TPL trial studied here is described first and foremost as personalized and data-driven. Looking into the historical and contemporary context of personalization we see how the idiographic or N-of-1 approach is in fact more similar to the QS movement and the personalization at work in recommender systems than precision medicine to which it is oft-compared. This insight, while alone is relevant for understanding how computer science and psychiatry converge at the field site, also lays the groundwork for chapters 2 and 3, which explore the theory of the empowered individual at the TPL.

The data-driven nature of the trial is the result of a historical trajectory in psychiatry that goes back to the decline of psychoanalysis. This trajectory is characterized by a devaluation of

questions of meaning and interpretation in favor of a more prosaic data-driven way of understanding patients. Chapter 3 will explore the consequences of this shift away from theories of meaning, which constitute a powerful tool for psychoanalysts who use it to explain contradictions and disjunctions in their diagnostic work.

Having established this historical shift and its consequences for digital psychiatry today, the interdisciplinary call to rectify equity issues in ML-based medical interventions through improved systems for explanation and interpretation appears not as an incidental oversight or technical limitation, but rather as a contradiction of the data-driven digital health approach. These issues cannot be addressed by simply adding explanations post-hoc. Therefore it is essential to better understand how practitioners conceptualize their work when it is broadly described as data-driven and personalized, and the ways this conceptualization affects how new kinds of work are accounted for (Chapter 2) and the role of the participant is expected to play in the digital mental health intervention (Chapter 3).

Chapter 2 Reconciling Healthism and Techno-solutionism in Formal Accounts at the Translational Psychiatry Trial

1. Introduction

This chapter describes the ways laboratory members describe their work, following the general approach of the dissertation which consists of comparing the theories laboratory members have of the trial to the observable practices that constitute it. Observational material is presented to show the theoretical constructs through which laboratory members describe their work in preparatory meetings and public discourses including presentations. Then, observational material from PWP guidance meetings is presented to show the interactional practices that constitute the trial as I observed it.

In advance of the trial, and in general descriptions of it as it got underway, the trial was characterized as data-driven, evidence-based and highly personalized—a way of helping participants understand themselves through data and thus learn new strategies for improving their mood. By this account, the outcome of the intervention would be determined by the combination of two primary factors: the participants themselves, who are characterized as capable and intrinsically motivated; and the machine learning system, which would empower participants to achieve their mental health potential.

The intrinsically motivated participant on one hand, and the insightful ML system on the other, correspond to the two primary disciplinary approaches in the lab: the psychiatric and the computational. The interdisciplinary collaboration of the trial therefore includes an effort to reconcile these approaches, and develop a complementary dynamic between the agential, empowered participant and the insightful ML system. As is explained below, cohering the two sides of this conceptualization of the digital mental health intervention requires continual discursive work, in the form of abstract descriptions of scientific work that address and repair contradictions. This discursive task is distinguished from the interactional work of, for example, training guides and guiding participants, which are analyzed as practices rather than discourses (both below in Section 5 and in later chapters).

The capable participant and the empowering data system each correspond to a sociological theory of science and technology: health discourses that emphasize the responsible individual have been discussed by sociologists under the banner of "healthism," and those that focus on the promising epistemological power of new technology are often called "techno-solutionism." These critical perspectives provide a counterpart, in the realm of social theory, to the two parts of the trial as it is presented in formal accounts in the laboratory. By first analyzing the conceptualization of the trial as it is maintained in formal accounts using these two sociological theories, and then offering an account of observed lab practices (later in this chapter and as the primary content of chapters 4 and 5), the present dissertation shows how formal descriptions fall short of adequately accounting for important kinds of work, and also how they influence this work itself. Specifically, the discourse by which the empowered patient and the insightful technology are reconciled has consequences for the ways labor is accounted for.

In formal descriptions of the trial, guides are rarely mentioned and when they are, their role is made to seem mediatory. Observation of the course of the PWP guidance makes clear that the guide is an important component of the intervention, and that the PWP guidance process includes forms of care that complicate the account of the trial which focuses solely on the participant's behaviors and their knowledge of data regarding those behaviors. The guidance work is a form of what Oudshoorn (2008) calls "invisible work" which serves to repair the gaps and inconsistencies that inevitably emerge as data science projects meet the contingencies of everyday life (Pink et al., 2018). The failure to adequately account for this work follows from the focus on participants and technologies in formal accounts of the trial.

In accounting for the work of guides, lab members relegate guidance duties to the realm of the technical, describing them as supplementary to technologies which, in their fullest and most reliable instantiation, could function without the need for guides. The guides themselves also conceptualize their work in this way when they describe it. Guides structure their work around maintaining their status as technical support and avoid infringing on the agential capacities of the participant, who is supposedly empowered by the ML system and the analyses that it has produced.

The characterizations of care work as supplementary and technical, and the ways it was modified to better align with these characterizations relates to the methodology of the present approach. Attentive listening, presence, and engagement with participants' affect and material circumstances are practices that the present observation approach and PWP guidance share. Studying how care practices like those of PWP guidance are integrated into health interventions that are formally described as computational, sheds light on the challenge of applying

observational methods to computational projects generally, and raises questions about remote, digital observational studies like this one as they relate to traditional ethnography.

2. Literature review

One common strand of critical social science work on this topic critiques the way digital phenotyping contributes to a trend of healthcare discourses in which "the responsibility lies primarily…with the purportedly empowered patient," minimizing the role of structural and environmental factors (Pickersgill, 2019, p. 21; see also Lupton, 2013a, p. 395). Approaches that subscribe to this individualistic view are termed "healthism" (Crawford, 1980). This line of critical inquiry is particularly relevant to the present study as it addresses an emphasis on individual participants as responsible and capable masters of their own health—critical sociological literature on healthism corresponds to the participant side of the two-sided frame through which the TPL conceptualizes their trial.

Just as there exists a social science literature on the concept of an empowered, responsible participant, there is also a significant literature discussing the other side of the digital health framework observed in formal accounts at the TPL. The wearable technology and the machine learning algorithm that makes use of its measurements are discussed as empowering patients in an objective, unbiased way in promissory discourses that have been termed techno-utopian or techno-solutionist (Lupton, 2014; Pickersgill, 2011, 2019). The present study, particularly our analysis of formal accounts of the trial, demonstrates how healthism and promissory, utopian and solutionist technology discourses are made compatible in the digital psychiatry context. We then compare this discourse, which maintains cohesion between the theory of the responsible participant and the techno-solutionist theory of the transformative healthcare AI technology, to an observational account of PWP guidance practices. By understanding the framework through which laboratory members describe the trial in formal

accounts, we can better understand why some labor practices do not fit into this framework in their full complexity.

Accounts of digital health programs as the accomplishment of either a new technology or a new kind of empowered patient both relate to the notion of human and nonhuman agency, which is a major topic in sociology and science and technology studies (STS). Sociologists have emphasized the way new health technologies take on agency by disciplining and controlling their users, exerting on them new responsibilities that, through the healthism discourse, are made to seem like empowering features (U. Beck et al., 1994). The present study, following Pols et al, acknowledges healthism as a useful way of understanding the history of health technology, but with the potential limitation of "[imagining] a too-uniform ('neoliberal') subject" (Pols et al., 2019, p. 99). Instead of describing how participants are disciplined by nonhuman technological agents, this chapter shows that, while formal accounts may follow this logic of healthism, lab practices constitute both gaps in this account and efforts to repair these gaps (Pink et al., 2018).

The present chapter thus aligns with Barad's agential realism approach, which understands boundaries and agencies as fundamentally open-ended components of sociotechnical systems that are always rearranging themselves in practice, constituting themselves as hybrids of bodies and machines, representations and movements (2007). The present study does not take its primary agents for granted—it includes a critical comparison of formal accounts, which focus on the agencies of the participant and the ML system, and observed care practices that struggle to adhere to the boundaries made in formal accounts. This approach is particularly powerful in the present context as it does not assume a commonplace separation between calculating machinery and caring human attention, but rather directs us toward the combinations and intersections that challenge this dichotomy (Mol et al., 2010b; Semel, 2021).

Marent and Henwood elaborate this method as the "sociomaterial approach," which distinguishes itself by close attention to "how algorithmic systems come to work in medical practice" (2023, p. 47). This approach differs from the substantialist theory that underpins technological determinism, which yields accounts of new digital health projects exclusively in terms of new objects and the physical and electrical connections between these objects. Instead, the present study is open to interpreting how objects and agents come to be seen as such in a dynamic interplay with embodied users, developers and scientists who bring their own conceptual and communicative schemas.

The problem of minimized, invisible or underacknowledged labor practices brought about by digital health initiatives has also been addressed in the existing sociology literature. In what Lisanne Bainbridge calls an "irony of automation," new sociotechnical systems that are promised to replace human workers often end up recruiting new, more specialized forms of human labor in case the system needs repair or maintenance (1983). Digital health is one of several contemporary fields in which projects that emphasize the role of new AI, ML or automation technologies downplay the forms of human labor that also help constitute the new sociotechnical system (Gray & Suri, 2019). Invisible labor often serves to repair gaps or inadequacies in data systems that require human work to be enacted in real situations (Mol et al., 2010a; Pink et al., 2018; Schwennesen, 2019). In Nelly Oudshoorn's study of patient self-diagnosis through telemonitoring, for example, individual motivation and responsibility is emphasized in formal accounts, following the logic of healthism, and the care work of home nurses is thus made invisible (2008). The concept of invisible labor in digital health has also been applied to the production, completion, validation and sorting of data sets in data-intensive clinical settings (Bossen et al., 2019; Green et al., 2022).

3. Approach

Following the general approach outlined in the introduction to this dissertation, this chapter consists of observation of laboratory practices including presentations, meetings, trainings, and PWP guidance sessions. I took field notes during observation and, when possible, digitally recorded teleconferences for later transcription and analysis. I also downloaded a variety of technical documents including datasets, spreadsheets, and PWP guidance worksheets. These documents provide necessary context for some lab interactions. I followed the trace ethnography approach, collecting observational data and also technical traces, using one to guide my exploration and interpretation of the other (Geiger & Ribes, 2011). As is discussed in further detail below (Section 6), the remote, digital, nature of this observational approach distinguishes it from what is traditionally considered ethnography.

Observational data including video recordings, transcriptions, field notes and digital traces were analyzed through a process of thematic tagging, excerpt selection and then interaction analysis. Firstly, data were tagged for themes of interest, following recent work in social science of digital mental health (Algera, 2023). In this case, themes cover topics like intrinsic motivation of participants, the importance of personalization, and chatbots as potential replacements for guides. This process of thematic analysis is designed to produce a sense of how the trial is formally accounted for in advance of its deployment. These accounts are presented as quotations in the text.

Also of interest to the present study are the instances in which these accounts are contested or modified as the lab initiates the trial. These instances were selected for further analysis as excerpts (Maslen & Harris, 2021). Observational data of these moments of practical coping and repair, including PWP guidance sessions themselves, are subjected to an interaction analysis based on Jefferson transcription (Jefferson, 2004; Sacks et al., 1978). This approach

follows from the below argument that formal accounts become complicated not by formal or theoretical contradictions, but as a result of the contingencies that arise as the actual doing of the trial takes place. The authority of what I call the formal account is, like all human activity, also accomplished through particular interactions. For an analysis of the interactional process by which PIs establish authoritative accounts as they are empirically realized (rather than its structure as a discourse) see Chapter 4.

Because of the disembodied nature of computational regimes, qualitative, interpretive, observational and ethnographic methods might seem inappropriate or out of place. In the present case, however, presence and engagement, which characterize important components of the ethnographic method, also constitute a part of the intervention, even with its emphasis on ML and wearable technology. While these aspects of the intervention are downplayed in formal accounts, observational methods are especially well-suited to bring them to the fore, revealing unacknowledged care practices at work in the digital health intervention (Beaulieu, 2017). While the role of the body is restricted given the remote nature of the observed meetings, it nonetheless plays a role as PWP guides and their participants engage one another with talk and gesture. Close attention is paid to bodily movements and other details of observed sessions. Transcription methods from the field of conversation analysis are employed for their suitability for conveying these features.

4. Two agencies of the trial in formal accounts: the engaged participant and the ML system

In formal accounts of the trial, the primary factors involved in bringing about the desired outcome of the intervention are the participants themselves, who are empowered by personalized ML-based insights, and the data system, which produces new knowledge in a rigorous and objective manner. First, I will show how this conceptualization is established in the lab. Then I will analyze an instance in which a conflict emerges that demonstrates the work required to

continually maintain the theory of the digitally engaged participant and the empowering ML system.

In lab meetings leading up to the trial and PWP guidance trainings in which lab members learn how to guide participants through the trial, the participant is described as playing a fundamental role in the success of the intervention. The personalization aspect of the trial is said to serve the purpose of bringing out intrinsic motivation—a lack of intrinsic buy-in by the participant "defeats the purpose of personalization." The data system, the other important component in this conceptualization of the trial, serves to empower participants to make decisions that they tailor to their own lives. In this way, participants themselves are responsible for many of the implementational challenges of the trial: "What you're doing as a coach is just trying to highlight, here's an area, or here's, here's something that we could tackle, here's a, here's a, here is something that we could try to address. But they're coming up with a solution that works for them that's relevant for them and their family situation, with their lifestyle, with their ability to get fruits and grains and vegetables, and things like that." Structural factors like the availability of fresh food in a participant's neighborhood, and social factors like the need to care for family members, are abstracted from their broader contexts and said to be resolvable through the individual choices of the participant who is empowered to do so by the flexible nature of the recommended plans.

Along with this powerful role ascribed to the participant, the ML system is also described as a primary component of the trial, including in one formal account in which the PI claims that "the future is in tech for mental wellness." By this account, the ability to understand mood in terms of an underlying pattern of behavior that is revealed through the ML technology provides an essential form of insight that empowers the individual to make positive changes: "You tell

me, um, make me feel better, you know. How do I quantify that? Like my happiness itself is not so well quantified. I was like 'okay well don't think about your happiness let's just try to do this,' you know, goal of this much activity or this kind of diet or this particular sleep. So you move away and you suddenly feel like if that is the underlying cause, it is going to create a sustainable change…I think, I think the data here, to us, feels like it provides insight and sort of an empowerment." The two components, the individual participant and insightful data system that empowers that participant, combine to constitute the force of the intervention as it is described in formal accounts.

The former component, the participants themselves, falls into the domain of psychiatric expertise which includes theories of motivation, mood and well-being that are specific to the experience of the individual. The latter component, the machine learning system, falls into the domain of computer science and computational modelling. Developing formal discourses to describe the power of the computational psychiatry intervention therefore includes an effort to bring these perspectives together in a multidisciplinary conceptualization of mental health treatment.

The two PIs, however, one with a background in psychiatry and the other with expertise in neuroscience and computational modeling, do not entirely agree on the balance of the two key components. In one participant's case, transcribed as Excerpt 1, the algorithm found that variables relating to both diet and social connection had a relationship to mood. When the two PIs are preparing the PWP guide by picking a plan and instructing them on how to propose the plan to the participant, they come to a disagreement. This disagreement relates to the question of the comparative importance of data insights and intrinsic participant motivation. PI1, who is trained in neuroscience and has expertise in ML and computational methods, suggests that if they

offer a diet plan to the participant then it won't qualify as "data-driven" (line 18). To this the psychiatrist, PI2, responds by suggesting that the data serves to elicit motivation to the participant, comparing the data science component of the

Excerpt 1. PIs discuss the comparative importance of the data-driven basis of the suggested plan and participant willingness to be enrolled in the plan.

```
1    PI1  Essentially, I think what we're trying here is to see (.) uh what a
2         database recommendation (.) A, would Participant be on board with
3         that, and then also if um that works in any way. I guess we're not
4         trying to put in our biases
5                                                    ((turns head to face PI2))
6         On top of that (.) also so
7         ((raises eyebrows))
8         (1.5)
9    PI2  We're not trying to (.)  yeah (.)
10                 ((turns to face PI1))
11        Well, well what would be (.) our bias? (0.8)
12                          ((turns head forward to face camera))
13
14        what do you mean by that?
15         ((turns head forward to face camera))
16   PI1  Um (.) I'm, I would say that if we were to recommend other plans it
17        would not necessarily… like the diet plan I would think is not
18        necessarily data-driven ((uses cursor to select part of bar graph
19        shown on shared screen))
20        it's just um
21        ((turns head to face PI2)) OK the participants wants ((noise))
22        necessarily data driven plan. ((turns head forward to face camera))
23        which is, which is fine (.) uh we….
24   PI1           ((turns to face PI2))
25   PI2           ((turns head away from PI1 to the left of the camera))
26        well I think it's part of the idea that(.)somebody has to still buy
27        into it
28                               ((turns head right to face camera and then PI1))
29   PI1                              ((nods and turns head to face forward))
30   PI2  and if they don't buy into the recommended plan, what plan will they
31        buy into?
```

```
32  PI1  [[Yeah (.)Agree.]]
33
34       Sure
35
36  PI2  I mean. It's like psychotherapy research. If you randomly assign
37       people versus if you like (.) oh which type of psychotherapy would you
38       like people generally do better if they're allowed to choose the type
39       of psychotherapy you know. It's a little bit like that. It's in the
40       same direction
41  PI1                                    ((turns head to face forward))
42  PI2  I think is what we're trying to do is like (0.2) here's our thought
43       about what might be the best for you but you know, do you buy into
44       that?
45  PI1                                   ((raises eyebrows))
46  PI2  If you don't, which one do you buy into that still may be the second
47       best
48  PI1  Right (.) sure
49  PI2  At some level, also this is just me philosophizing, but I think part
50       of the motivation for data  is to get people to buy into it.
51       ((turns head to face PI1))
52       If they think oh I'm, I'm likely to benefit from this, they're more
53       likely to benefit. It's like, you know, you're like kind of massaging
54       the placebo effect. You're massaging, like, kind of, behavioral
55       change. So, so the buy-in may be the most important thing but the data
56       is helping with that.
```

intervention to a placebo pill (line 54). If the participant thinks they are operating on the basis of a statistically relevant relationship between activity and mood, then they are more likely to succeed in their efforts to change their lifestyle. For PI1, this conflicts with the notion that the choice of plan, because it is made in a data-driven manner, is free of biases of the investigators.

The possibility of observing this kind of disunity in perspectives within the lab is a unique affordance of the observational approach. Formal accounts from this lab, as is typical, tend to present lab views as uniform and cohesive. The conflictual nature of this exchange is made evident largely through bodily movements. Because the PIs are both using the same

camera to participate in the ZOOM meeting, they are able to face one another to engage in a dialogue in which the trainee, who is participating remotely through their own camera and computer, is never addressed directly. The fact that the PIs engage each other in this discussion during a guidance training session demonstrates how implementational processes like trainings can manifest important disagreements that did not come to the fore during the trial design process. In this case, the ambiguity of the ML output, and the subsequent process of picking a plan from an unclear ML recommendation, gives rise to an exchange that reveals a fundamental difference in approach between the PIs.

Disjunction between the computational and psychiatric approaches is not uncommon in digital mental health initiatives. In a proposal for a framework for ethics and accountability in machine learning for mental health, McCradden et al. elaborate the notion of the "AI chasm" between better technologies and better outcomes in the clinic, as "epistemic (methodological) and ethical, generated by a clash between the (epistemic and ethical) cultures of computer science and clinical science (Wongvibulsin and Zeger 2020)," suggesting that "if AI is to be translated successfully into the clinic, the values and methods of these disparate cultures must be reconciled" (2022, p. 8). The authors attribute this disparity to differences in methodology, including approaches to validation which, in the ML context takes on a strictly technical, algorithmic meaning, lacking consideration of contextual factors that have come to be seen as salient as clinical research ethics have progressed historically. The excerpt analyzed here, however, highlights another difference between the two approaches: the primary agents seen as responsible for bringing about clinical outcomes.

In an article on the hopeful and promissory discourses around treating psychopathy using novel insights from neuroscience, Martyn Pickersgill traces the establishment of epistemic

authority by neuroscientists and their findings, quoting a clinical psychologist who locates this authority in a concept of agency: "I suppose [neuroscientific knowledge,] for me raises big challenges about what self-control and self-regulation is, and to what extent are any of us in control of our behaviour, and to what extent our behaviour's controlled by processes of which we are aware or conscious. And I suppose neuroscience is challenging that, you know those notions about free will" (Dr. 2 qtd. in Pickersgill, 2011, pp. 458–459). In the case of the clinical psychologists of psychopathy that Pickersgill studies, the individual will of the participant is called into question by neurobiological mechanisms, the revelation of which became the basis of epistemic authority of neuroscience over psychiatry. As a result, the scientists and clinicians Pickersgill observes are said to conceptualize their work through the lens of "techno-somatism," which sees the body, affect and emotion as analytically comparable to the circuits and chemical mechanisms of neurobiology. In a related discussion of a digital health intervention based on monitoring of heart-rate variability, Minna Ruckenstein shows how a reconfiguration of agency lead to a similar kind of epistemic authority, and even disciplinary power, this time endowed not to neuroscience but data science, as digital health application users not only felt influenced by the analytics and recommendations of their ML based applications and devices, "but they expected the data flows to act on them" (2014a, pp. 73–74).

Rather than subordinate the participants to the force of the data, in our case the PIs seek a balance between the agential participant and the insightful, empowering algorithm. Efforts at reconciling these approaches in ordinary laboratory discourse speak to the prospect of reconciling the broader ideologies of healthism and techno-solutionism by bringing the individual and the algorithm into a complementary theory of agency. The computational approach includes an emphasis on the power and promise of new ML technologies that has been

termed techno-solutionism or techno-utopianism (Lupton, 2014; Mills & Hilberg, 2020). The psychiatric approach includes an emphasis on the potential for the participant to bring about their own recovery by activating their full potential, termed "healthism."

The balance between the structuring insights of the ML system and what Lupton calls "the digitally engaged patient," who needs only to responsibly and prudently activate their inner potential, reflects an essential problematic of neoliberalism (Lupton, 2013b). This tension has also been observed in ethnographic work on the use of wearable devices for self-care—devices which offer a "nudge" to remind users of a behavior they should take or avoid: "the nudge is a curious mechanism, for it both presupposes and pushes against freedom; it assumes a choosing subject, but one who is constitutionally ill equipped to make rational, healthy choices" (Schüll, 2016, p. 12). While this contradictory conceptualization of power and agency, which has been termed by philosophers "libertarian paternalism"(Thaler & Sunstein, 2008), might seem like a new inversion of neoliberal discipline, it is rather an extension of the paradox of a freedom that, to be sustained, must also be continually adjusted through discipline, which is at the heart of the historical emergence of neoliberalism (Foucault, 2008).

Excerpt 1 demonstrates not only the challenge of bringing together two scientific fields, but also developing a theory of agency that maintains coherence between the two sides of the trial as it is conceptualized in formal accounts: the empowered participant of healthism and the insightful technology of techno-solutionism. Bias is mentioned as the first argument for prioritizing the data system over the intrinsic potential of the participant. In response, the placebo effect is evoked as a potentially valuable therapeutic avenue, regardless of its statistical validity. The placebo effect aligns with the notion, common in healthism, that participants have the intrinsic motivation to change within them, and therefore can take on responsibility for their own

health outcomes. The promise of bias-free, objective insights, on the other hand, is typical of promissory discourses around new health technology. The disagreement between the computational and psychiatric perspectives, then, is also a challenge to the coherence of techno-solutionism and healthism.

5. The minimal role left for PWP guides

The role of PWP guides who meet with participants to help them implement the ML-based recommendations follows from the conceptualization of the trial as consisting principally in the self-improvements of participants which are enabled by the insights of the ML system. Formal accounts of the PWP guides constitute part of the discourse described above through which the uneasy cohesion between healthism and techno-solutionism is maintained. The PWP guidance practices themselves, as they are observed, further demonstrate the challenges of maintaining this formulation, as they are influenced by it and also evidence its impossibility. As was done above, I will start by showing how the relegation of guides to the technical side of the intervention is established in general descriptions, and then look at practical instances in which this logic is called into question and needs to be repaired.

5.1 Relegation to the technical in the planned role of PWP guides

The guidance component is downplayed in the initial presentation of the trial. The PI suggests that the PWP guide could be feasibly replaced with a chatbot, and that the role of the guide is too minimal to be considered therapy as it is not the guidance but rather the participant's own behavior that constitutes the intervention. Later on, when presenting a revised, final version of the protocol to members of the lab, one of the PIs offers the following explanation of the role of the guide:

> It won't be, you know, completely unguided. You will get the data chart. You will get values of what you want to guide in the range, what range you want to guide in….. It won't be unsupervised at all. Won't be that, "OK hey you're getting the sleep plan, now go do it."

> So the first day is more about getting them um acquainted with what their data says and getting them on board with, "hey, this looks like this plan is something we want to do with you and our goals are, to see that somehow, to have you in the health behavior that we've chosen for you, or the plan has, or the data has indicated for you as the best plan."

PWP guides are reassured that their role will be subordinated to the insights of the ML system, which they will pass along to participants (for a deeper exploration of how this takes place in interaction included through use of data visualizations, see Chapter 5).

Of particular interest is the use of the phrase "unsupervised" as the PI reassures the guides of the structuring role the data will play. Supervised and unsupervised are terms used to describe ML systems. The former category make classifications based on a set of training data that has already been labeled by a human. The latter determine their own labels *de novo* by discerning patterns in the main dataset. By using the phrase "supervised" to refer to the human interpreter of the ML output, the PI reverses the typical meaning of the term, subordinating the human to the algorithm. More generally, this language integrates the human guides into the technical apparatus of the computational system, according to which some ML models are supervised by training datasets, and some PWP guides are supervised by ML outputs. The technical nature of PWP guidance, as it is conceptualized in advance, is also reflected in the lack of clinical training required of guides. Lab PIs encourage any lab member to join PWP guidance trainings, assuring them that the work is not a form of therapy.

This description of the role of the guide aligns with the broader conceptualization of the trial—guidance work is part of the technological component of the intervention. By observing these practices, it becomes clear how they exceed the formulations described here, but also how they are affected by them. The relegation of the guides to the technical side of trial goes beyond

the discursive framing we see in formal accounts, influencing the way guides consider and carry out their work.

5.2 Relegation to the technical in guidance practice

Ethnographic attention to PWP guidance sessions reveals how this practice exceeds a transparent transmission of data insights to participants. Far from arising through participants' own intrinsic motivation, the treatment plans that are assigned on the basis of the ML categorizations are carried out with the support of the PWP guide. While the guide's role was minimized in planning sessions and presentations leading up to the trial, in PWP guidance sessions themselves, the guides position themselves as present and engaged in ways the data system itself, even if augmented by a chatbot system, could not likely achieve.

The presence and engagement of the PWP guides often takes the form of listening. In the case of one participant engaged in the positivity plan, the second session includes the participant sharing extensive personal details about their goals and perceived shortcomings. This is, in a sense, merely part of helping the participant complete the necessary handouts—another evidence-based and supposedly independent component of the plan. Reading back one's goals and fears to a guide who is listening attentively is not an independent activity, however, regardless of the supposed passivity of the PWP guide in this case.

Excerpt 2. Participant questions whether reflecting on a tough week is part of the intervention.

1	Guide	How was last week?
2	Participant	U::h, last week was kind of rough (for me), just like,
3		personally. But I wasn't feeling especially grateful for
4		those (.) in those grateful worksheets=
5	Guide	=yeah=
6	Participant	=but I, but I fill(h)ed it out
7	Guide	So uh, would you like to tell me what's going on or
8		anything like that? Do you wanna talk about it at all?
9	Participant	I mean, uh, I don't know do you, is that a part of the
10		th(h)ing?
11	Guide	Uh, no, but, you know, just knowing that the, uh you know,
12		last week has been tough so that (.) your gratuity levels
13		are lower, you have found the more mundane things, uh,
14		positive in your day — uh that's, you know, kind of enough
15		for me to know. U:h, okay, let’s look at what you’ve got
16		then.

In Excerpt 2, the ambiguity of the PWP guide’s role leads to confusion for the participant who wonders about the limits of the kind of talk involved in the intervention. The PWP guide responds to this uncertainty in two ways. First, the guide suggests that they do not need to know details about what made the participant’s week feel like a “tough” one, because they already had surmised that the participants “gratuity levels” had been “low” (lines 12-13). Here the guide evokes the quantitative logic of the ML system on the basis of which the participant was assigned the positivity plan in the first place. The evocation of high and low levels harkens back to an early lab meeting in which the PI describes the objectives of the guidance practice: “The goals really are to stay in the zone of consistency for the plan that you’re recommending. So the exercises that you'll give to them as part of the sleep plan, or exercise plan in this case, would be that they can maintain that whatever levels of data helps them get that, um, you know, happy mood state.” This advice is given in the context of an example of a participant who had been

assigned the exercise plan, which is a more intuitive context for this kind of quantitative logic than that of positivity. Nonetheless, the PWP guide observed in Excerpt 2 forwards the emphasis on being in the correct "zone" or at the ideal "level" to resolve the challenge. Second, the guide redirects the conversation to the handouts with the prompt, "let's look at what you've got" (line 15). This move serves to transition from an unstructured conversation about the participant's recent affective experiences into the "evidence-based" domain of the handout. The handout also serves to limit the conversation to topics the participant already decided in advance were appropriate to share with the guide.

The move back to the quantitative conception of mood and the premade handouts follows from the PWP guide's training. Listening is explicitly ruled out as a role of the PWP guide in the first preparatory meeting held by the clinical team: "We're not really doing therapy per se. It's um, it's, we're not, you're not, you know, there to like really listen to everything that's going on with them. It's more that it's a very action oriented plan. All of these plans are actually evidence-based. They already exist in the literature and, and we're giving them out exactly how they've been cited in in the existing literature." The existing literature on which the handout is based is used in a clinical team meeting to explain how the PWP guidance stops short of therapy. The guide in Excerpt 2 uses the handout as a way of moving away from an unstructured question about the participant's week, which perhaps seemed too much like therapy, back to the evidence-based intervention.

The use of the handout and the quantitative language of "gratuity levels" serves to reconcile present, engaged listening as a part of the supposedly objective, ML and evidence based intervention. Another version of this technique is reported in Beth Semel's ethnography of a digital mental health project which aimed to develop a voice-based diagnostic system. In that

case, Semel observed researchers who, in their efforts to produce a training dataset of correctly labeled voice samples, attempted to "listen like a computer," ignoring semantic content and focusing only on labeling tones of voice (2021). The handouts are like the schematics with which researchers in Semel's study categorized tones of voice—both serve to convince the researchers and participants involved that the humans were standing in for an objective, mechanical system (Seaver, 2011; 2021, p. 17). In Semel's case they were developing a training data set so they could eventually be replaced by a ML system. Similarly, in the PWP guidance case it is suggested in formal accounts that a chatbot system could, in future versions of the trial, replace the PWP guides.

Semel directs us to consider the political consequences of humans aiming to imitate the mechanical objectivity of machines, including the way that seeing work as mechanical can have depoliticizing effects, obviating questions of what kinds of workers are recruited, how they relate to their participants, and structural forms of oppression in the context of which mental health problems arise (Tomičić et al., 2021). Semel also connects her study to the broader literature on invisible labor, pointing out how digital projects that minimize the role of humans, who are said to merely bridge the gaps between technologies, often times end up recruiting more human labor, not less (Oudshoorn, 2008). In Oudshoorn's discussion of invisible labor, she develops a typology of the kinds of labor that are made invisible in the digital health context. She describes "affective work" as "emotional support to create 'intimacy at a distance' "— Excerpt 2 shows an effort to conduct this kind of work (which also features prominently in Chapter 5), and the confusions and complications that follow, because this kind of work is not accounted for in formal accounts of the trial.

The particular discourse by which guidance labor is made to seem technical and easily replaceable is part of a larger conceptualization of the trial which is based on a fraught combination of the empowered participant of healthism and the insightful new technology of techno-solutionism. Observation demonstrates the complexity of PWP guidance, while attention to formal accounts demonstrates how this complexity becomes obscured and how the practice is affected by formal descriptions of it. Not only does the discourse around the empowered participant require on-going repair in laboratory meetings, as is shown in Section 4, but also reconciling these descriptions with the real practice of PWP guidance requires continual negotiation of the kind demonstrated in Excerpt 2.

6. Conclusion and methodological reflection

The present chapter demonstrates how the TPL trial is described by lab members: based largely on the efforts of a participant who is empowered by analytical insights of a personalized ML system. The conceptualization includes a primary role for the responsible individual participant—a role often emphasized in discourses that have been termed "healthism." It also includes a primary role for the computational system, as is common in promissory discourses around new technologies. Balancing the agential capacities of these components requires discursive work in laboratory discussions and the contradictions it yields require repair. These findings show the strength of the sociomaterialism approach, which lacks an *a priori* notion of relevant agents and instead looks to laboratory practices to show the dynamic interplay of human and nonhuman components. The present analysis also complicates the notion of the neoliberal subject who is disciplined by systems that are said to empower them. The emphasis on the individual is not a constant, dominant discourse, but rather part of a negotiation.

As part of the formal account of the trial, PWP guidance work is relegated to the technical, which affects the way guides describe and carry out their work. These findings

elaborate the processes by which new forms of human labor can be made to seem like unimportant or invisible parts of digital health interventions (Gray & Suri, 2019; Oudshoorn, 2008). Also, the manner in which the care work of PWP guidance is conceptualized as part of the technology-focused intervention has implications for how social scientists conceptualize their own qualitative and ethnographic inquiries into projects that are described as computational.

The observational approach used here follows the example of remote, digital ethnographies and leverages technical traces following the "trace ethnography" method (Geiger & Ribes, 2011; Lupton & Watson, 2021; A. Watson & Lupton, 2022). However, it also lacks kinds of involvement that distinguish it from ethnography as it is traditionally defined. For example, the approach does not access the full spatial and sensory context in which participants learn to use wearable devices and complete activities for the intervention. The lack of physical presence also has implications for the degree to which the observer is implicated in laboratory practices, as a ZOOM observer does not occupy laboratory space or participate in the embodied, spatial dynamics of in-person laboratory work.

The remote nature of the observations also affords new insights into remote scientific work and telehealth, because it reflects the situation of PWP guidance training and PWP guidance itself. While the approach lacks embodied, emplaced co-present interaction, it nonetheless includes interactional aspects that link it, and PWP guidance, to ethnography. Anne Beaulieu describes how the attentive listening of approaches like PWP guidance and ethnography can come into conflict with computational regimes:

> In terms of a method where the "ethnographer is the instrument," some of the central claims of computational regimes pose a radical challenge. Presence and engagement, two key elements of ethnographic approaches, have been formulated according to the liberal humanist tradition that links the ethnographer as investigating, learning and knowing

> subject to embodiment. In a computational regime, where informational patterns are privileged over material instantiations, bodies matter less and less (Hayles 1999), therefore posing a challenge to ethnographic traditions. (2017, p. 30)

Rather than simply a less rich version of ethnography, the present method can be seen as a hybrid of kinds of practices that are often kept analytically separate: attentive listening and present engagement on one hand, and computational, quantitative data science on the other. Coping with the lack of embodied co-presence, as has been investigated by sociologists of telehealth (see eg. Grosjean et al., 2021; Mort et al., 2003), is a challenge that is common to the remote social scientist and the remote PWP guide or clinician.

The present analysis of the uneasy relationship between PWP guidance and the other components of the intervention, including its characterization in formal accounts, also sheds light on how ethnographers build and sustain relationships around the computational projects that they study, addressing questions raised by Bemme et al. : "How might anthropologists and psychiatrists be 'cousins' when it comes to new forms of tracking, tracing, and making sense of mental (ill) health? How does the digital psychiatric subject relate to the ethnographic subject? What are 'we' representing?" (2020). The challenge of adequately accounting for the labor of PWP guides speaks to the difficulty in reconciling the ethnographic and the computational. The important work that PWP guides conduct as part of the trial, however, also shows the relevance of engaged, present, embodied and interpretive approaches even in computational psychiatric interventions, and the suitability of ethnographic social sciences for investigating these components which can be made to seem comparatively unimportant in formal accounts.

Beyond sociological methods and theories, the present chapter also contributes to discussions around study design and computational system design. The interactions of PWP guidance sessions include, for example, reflections on "lived data"—the ways in which the data

system is entangled in everyday life, which Sanches et al aim to capture in their proposal for a design protocol informed by Barad's theory of agency (2022). The interactions observed for the present study also include moments of contestation. "Contestability" has been suggested as an important design consideration for ML based systems for psychotherapy, as opportunities for contestation allow for designs that accommodate shifts in relationships between people, professions and technologies (Hirsch et al., 2017).

Creating opportunities for contestation and critical reassessment of the design of a study and the roles of the laboratory members involved follows from the suggestion of Nete Schwennesen and other sociologists of digital health who call for "response-ability"—a framework for modifying the material conditions and observable practices of clinical work—as opposed to just transparency (Dudhwala & Woolgar, 2019; Schwennesen, 2019). One kind of response the present study supports is exemplified by recent moves toward co-design for health technology (see eg. Sanz et al., 2021; Wheeler et al., 2022) by demonstrating the value of attending to participant–guide discourses on the implementation of PWPs. While proposals for co-design often emphasize the role of participant or patient feedback in the design process, the present study demonstrates the importance of thoroughly specifying the guidance, coaching and other interactional roles required to elicit this feedback.

Acknowledgements

Chapter 2, in part, is a reprint of the material as it appears in Berners-Lee, B. (2023). Reconciling healthism and techno-solutionism: An observational study of a digital mental health trial. Sociology of Health & Illness, 1467-9566.13683. https://doi.org/10.1111/1467-9566.13683. The dissertation author was the sole and primary investigator and author of this paper.

Chapter 3 Care Practices and Environmental Factors Behind the Multiple Selves of the Digital Psychiatry Participant

"Formerly we were haunted by the fear of resembling others, of losing ourselves in a crowd; afraid of conformity, and obsessed with difference…All that matters now is only to resemble oneself, to find oneself everywhere, multiplied but loyal to one's personal formula….Resemblance is no longer concerned with others, but rather with the individual in his vague resemblance to himself; a resemblance born of the individual's reduction to his simple elements" (Baudrillard, 2008).

1. Introduction

The previous chapter discussed how formal accounts of the trial emphasize two components as primarily responsible for bringing out changes in mental health: the ML-based technology and the participant. PWP guides, who offer participants help interpreting and implementing behavior plans, attempt to align themselves with the technological side of this formulation by evoking the language of machine learning, statistics and data science when discussing participant mood apart from the digital measurements of it. In this chapter, we investigate the participant side of the technology-participant dyad, showing how, in formal accounts, the participant is multiplied into many versions or layers to accommodate the

heterogenous relations participants have with technologies, lab members, and data, as well as the environmental contingencies involved in a participant's behaviors as they change over time.

In preparations for the TPL trial it was characterized with an emphasis on the personalized nature of the intervention. Personalization was to be facilitated by the multiple options of ML algorithms that could be used depending on which would create the most powerful model of a particular participant's behavior and mood, and the multiple PWP options. The system was described as a way of helping the participants understand themselves through data and thus learn new strategies for improving their mood. The "PWP guides" were said to be replaceable with chatbots, and their role was made to seem technical and mediatory. As the guides were trained, they learned a theory of the participants that they would be communicating with. According to this theory, participants have internal motivation to make positive changes to their activities (eg. eating, sleeping, walking, exercising), and this motivation can be elicited by showing participants their data—specifically, the analysis of their behavioral data as they relate to their reports of mood. This theory of the participant accords with the minimal, technical role of the guide. All the guide is required to do is show the participant the relationship between their behaviors and their mood and the participant will then experience an intrinsic desire to act on the basis of this connection by modifying their behavior.

The central contention of this chapter is that multiplication of the participant into many "layers" or "inner selves" serves a similar epistemological role as the notion, common in poststructuralist theories of language and communication, of single entities containing or imparting multiple, sometimes contradictory meanings. The philosophical pragmatism of Richard Rorty is used below to show the parallels between multiplying meaning and multiplying entities as convenient reductions for describing complex systems with heterogenous attributes

and relations. As was established in Chapter 1, this connection accords not only with the recent history of communication theory, but also of psychiatry, in which the paradoxes of meaning at the heart of psychoanalytic psychotherapy have fallen away in favor of psychotherapeutic interventions based around the individual participant's behavior. The dearth of attention to issues of meaning and interpretation in the field of digital mental health is not an incidental oversight, but rather a result of fundamental shifts in the way digital mental health practitioners conceptualize their work.

The observational approach and examples of ethnographic studies of scientific work are presented as an alternative to both theories of the individual who contains multiple meanings or the individual multiplied. This approach gets its comparative strength by welcoming the contingencies that emerge as individuals enact recommendations in the world, as well as the complex relations that individuals have with technologies, scientists, laboratory staff and data. This approach is better suited for producing robust descriptions of sociotechnical systems as it sheds light on the differences in how phenomena with the same name are measured in different contexts, and the practical challenges of implementing the ML-based intervention, which are often addressed with otherwise under-acknowledged and inadequately described forms of human labor.

2. Approach

The present approach, which is similar to that used in chapter 2, combines ethnographic attention to laboratory practices with analysis of descriptions that lab members give of these practices. Because I was able to observe planning for, training in and conduct of PWP guidance, I contrast descriptions of this practice that I hear in the TPL with my own description of the guidance as I observed it. As we saw in Chapter 2, this approach simultaneously foregrounds the empirical validity of observed interactions and practices while acknowledging that the formal concepts and

theories that scientists use to describe their work influence the course of laboratory activities. By analyzing the two side-by-side it is also possible to gain methodological insight into the strengths and weaknesses of observation, and to demonstrate how this approach can reveal what is otherwise left out when scientists describe their own projects. Analysis of laboratory accounts have been shown to productively supplement observations of laboratory practices, including in the digital health context (Berkhout & Zaheer, 2021; Pickersgill, 2019).

This chapter also makes use of the historical context established in Chapter 1 which, in part, forms the basis of formal descriptions and theories that psychiatrists and data scientists share. This historical context helps to explain the disjunction between descriptions of the trial, which focus on the multi-layered participant, and the trial as I observe it, which includes forms of labor and environmental contingencies that are left out of lab descriptions. The use of historical data alongside ethnographic observation is not uncommon in disciplines including urban studies, anthropology (Katz, 2010), and cognitive ethnography (Osbeck & Nersessian, 2006), where it offers a chance to understand how ethnographic subjects' conceptualizations of their activities build on collective memories and disciplinary histories.

3. The participant's multiple selves understand, act on, and feel the benefits of the digital mental health intervention

In lab meetings leading up to the trial, PWP guidance trainings and PWP guidance sessions themselves, the participant is described as an important component of the trial. The process that lab members hypothesize will lead to positive outcomes for participants is based on, first, the participant becoming aware of relationships between their behaviors and mood; second, the participant changing their behaviors according to these correlations; and finally, the participant experiencing improved mood itself. Each stage of this process, whereby the participant,

empowered by the ML system, brings about their own improvement in mental health, is described as the manifestation or revelation of a new part, layer or version of the participant.

The first phase of the intervention consists in the participant reading and interacting with ML outputs and thus coming to a new understanding about how their mood and behavior relate. In a PWP guidance training, one of the PIs suggests that participants, on some level, already understand these relationships: “The model shouldn't come as a surprise to them. The model should be like, hey, you know I don't know if you've noticed this but hopefully they have. But if they haven't, then we're providing some feedback using our, our machine learning model.” Knowledge of what kinds of activities cause different moods is already within the participant, such that showing them representations of this relationship is really making them conscious of something they previously only knew in a tacit sense. By this account, the participant and the ML model are epistemologically linked on the basis that both are logging moods and behaviors, each in their own way.

The notion of an inner knowledge of the links between mood and behavior continues in PWP guidance sessions. In his sessions with one participant, who had been instructed to complete a handout as part of the positivity plan, the PWP guide explained that “the purpose here is to realize your goals,” not in the sense of manifesting them in real life, but in the sense of bringing them to awareness. In order to realize the goals the participant may already have, the first step is to complete writing exercises in order to “have our goals in front of our eyes.” Writing, according to the guide, is one of the best ways for the participant to realize the thoughts that tend to hide within them: “writing is the ultimate way of thinking. If you think about it correctly, if you write something down, your brain is registering it twice. Once when they're thinking, twice when they're writing it down and a third time when you read it.” The reason for

the participant to read what they have written is to allow the conscious self to learn about the deeper self that did the writing.

The knowledge that participants find within themselves, however, is not only valuable for determining the behavior change plan, but also as a tool for unlocking intrinsic motivation to carry out this plan. In a PWP guidance training, the PI explains that the participant should have an inner understanding of what behaviors are correlated with their moods and, further, an inner motivation to use this knowledge to improve mood: “we're not the ones doing the work. We’re providing a frame or, or some kind of way to just ask questions. But really the individual, the participant is the one who's, like, really should be coming up with strategies or solutions.” As this excerpt implies, the choice of plan, in some cases when the data are not entirely determinate, can be put to the participant. In a discussion of one of these cases, also excerpted in Chapter 2, a PI explains how involving the participant in the choice of plan could help bring out intrinsic motivation which could be of clinical value: “At some level, also this is just me philosophizing, but I think part of the motivation for data is to get people to buy into it. If they think oh, I'm, I'm likely to benefit from this, they're more likely to benefit. It's like, you know, you're like kind of massaging the placebo effect. You're massaging, like, kind of, behavioral change. So, so the buy-in may be the most important thing but the data is helping with that.” Motivation is an important component of the intervention, but it must be activated by appropriate insights about how mood and behavior relate. The PI suggests that working with participants who lack intrinsic motivation “defeats the purpose of personalization”—the reason for giving plans that are accurately tailored to participants is because these better align with what they are intrinsically motivated to do. In this way, what the PI calls a kind of “placebo effect” is also an indication of a kind of validity, because for an insight to activate an inner sense of motivation, it must accord with the

participant's previously inaccessible inner sense of what activities improve their mood. In a comment from a PWP guidance session that suggests the importance of the participant's own potential, the guide explains that the trial "isn't a project that I can give you a pill or do something *on you* or anything like that, and *you* don't have to do it. It's a project that is really dependent on you as a participant, and whether you really buy in and want to do it or not. And you know, uh, if you're really on board it can give really good, you know, results."

The inner aspects of the participant include not only the knowledge of what actions would lead to a better mood and the motivation to take those actions, but also the better positive mood itself. This is communicated through an expression that is used in several sessions of the PWP guidance: positive emotions sometimes "show themselves," and at other times remain hidden. As was the case with the layer of the participant who already knows how behavior and mood relate, and the part of the participant that is intrinsically motivated to change, the participant that feels positive mood already exists at the start of the trial. Through the intervention, the participant becomes able to let hidden feelings of positivity emerge such that they can be consciously noticed and talked about.

There are three key moments in the supposed course to achieving improved mood: understanding which behaviors lead to better mood, making an effort to act out the behaviors that correlate to better mood, and finally experiencing the clinical outcome of improved mood. According to the theory expressed in trainings at the TPL, all of these key moments hinge on inner components of the participants themselves. This theory of the multi-layered participant sheds new light on the notion of the N-of-1, idiographic approach described Chapter 1. As is suggested by the epigraph of this chapter, the individual of the TPL is distinguished as such not

by standing out from a general population, but rather by differentiating itself from other parts of itself.

4. Multiple meanings, multiple ontologies, and the alternative presented by observation

Having shown how the theory of the multi-layered participant in established through formal accounts in which lab members describe the trial, we will now contextualize the theory as part of contemporary healthism and to show how it continues the historical trajectory of psychiatric theory described in Chapter 1. We will then explore the limitations of this theory, first through Richard Rorty's pragmatist critique of representationalism, which also applies to multiplying of entities or "ontologies." Observational and ethnographic studies of scientific work are proposed as an alternative because of their strength in accommodating multiple heterogenous forms of data and kinds of actors by looking at the local contexts in which these factors emerge, are measured and interact with observers and practitioners.

4.1 Theory of the multi-layered participant as part of healthism and technosolutionism

The theory of the multi-layered participant plays a role in cohering descriptions of the trial with broader discourses including healthism, which emphasizes the importance of the empowered individual patient in determining health outcomes; and also the dyadic formulation of digital mental health described in chapter 2, which sees participants and technologies as the actors that have the power to bring about changes in mental health. Participants at the TPL are described as responsible for three key steps in achieving improved mental health outcomes, exemplifying healthism—a kind of discourse characterized by an emphasis on individual responsibility for health, often at the cost of engagement with structural factors (Crawford, 1980; Lupton, 2013b; Schüll, 2016).

The discourse of the multi-layered self, which extends and modifies the responsible individual of healthism, also has consequences for how the other components of the trial are

described. By attributing many roles to the participant, the ML system (which sometimes seems to include the PWP guides, as is discussed in chapter 2) is left as a purely technical component that need not exert discipline on participants, as they already have motivation within them, and need not inspire new ideas in the minds of participants either, as these are already tacitly understood as the result of experience. This account is in conflict with previous ethnographic work which has shown how visualizations of data act on patients, making the data seem like a primary agent (Ruckenstein, 2014b, pp. 73–74). The account studied here, on the other hand, includes a theory of the multi-layered participant: data flows that might be said to "act on them" are actually just revealing to them another layer of the self. While the disciplinary force of the data is downplayed in lab accounts of the trial, the data nonetheless play a key role in revealing to the participant their own motivations and inner feelings.

4.2 Theory of the multi-layered participant in historical context: a development on theories of multiple meanings

The theory of the multi-layered participant that we hear in the TPL emerges from the recent history of psychiatry. As was explored in more depth in chapter 1, the TPL trial is based largely on cognitive behavioral therapy (CBT), and therefore, in formal accounts, is said to build on concepts from CBT including those that that distinguish it from psychoanalysis. The tradition of psychoanalytic psychotherapy, going back to Freud, includes central use of concepts around linguistic meaning. As it developed throughout the 20th century, psychiatric problems were treated by psychoanalysts as paradoxes of meaning that could be best resolved through interpretation and dialog.

While psychoanalytic psychotherapy eventually gave way to CBT as the dominant current in talk therapy in the United States, its theoretical orientation toward the paradoxes and vicissitudes of structures of linguistic meaning became the basis of new currents in

psychoanalytic philosophy and critical theory[14], which came under scrutiny by pragmatist philosophers. The work of Richard Rorty is of particular interest here because of the way he connects the problem of representational doubling or multiple meanings, to the problem of ontological doubling, or multiple entities or version of an entity. Through his work we can see how both forms of doubling can obscure the complexity of some entities and the heterogenous relations that link these entities differently in different contexts.

4.3 The pragmatist critique of representation as it applies to both multiplied meanings and multiplied ontologies

The philosophical critique of epistemology as a representational system that more or less accurately reflects the world is an important contention in 20th century pragmatism, which has made a lasting impact on sociology and science and technology studies. Richard Rorty locates the origin of a philosophy preoccupied with an epistemological "mirror of nature" in early modern thinkers, including John Locke. In his refutation of this position, Rorty suggests that we ought abandon "the notion of knowledge as accurate representation, made possible by special mental processes, and intelligible though a general theory of representation" (Rorty, 1979, p. 7). According to this theory of representation, entities exist in an elusive form and are only accessible to knowledge as representations. Therefore, the question of truth is a question of which representation most closely reflects reality. The epistemological challenge posed by these multiple, sometimes contradictory meanings or representations is a point of departure for post-structuralist theories of meaning, including some with a basis in the psychoanalytic framework of Freud and Lacan (see, for example, the poststructuralist hermeneutics of Paul Ricœur (1970, 2006)).

[14] This connection from psychoanalytic psychotherapy to critical theory is not to be confused with Schön's practice-oriented, intersubjective account of psychotherapy compared to the approach of the present dissertation in the Introduction. Schön's methods are primarily based in ethnography, not poststructuralist critical theory.

The pragmatist rejection of representationalism we read in Rorty, on the other hand, is a common thread that links several other prominent theorists of the late 20th century. The critique has resonated through fields like anthropology, communication and sociology, offering scholars the ability to forego mediatory questions of encoding and decoding representations and instead develop commentaries on how the material circumstances, people and things in the world directly act on one another and the observer. In this way, Rorty's pragmatism and its critique of representational and ontological doubling is theoretically coherent with recent ethnographies, especially those that emerge from the tradition of ethnomethodology (eg. Alač, 2020; Mol, 2002; Suchman, 2007). Leaving behind representation affords a new approach that takes things in the world as actively constituting it, including performative utterances, which act on people and things in the world rather than representing them (Austin, 1975), and an enactive mind which actively achieves the perceptual world rather than more or less accurately portraying it (Noë, 2004).

Rorty's explanation of the problems with representation, and the kind of epistemological inquiry that arises from it, is especially relevant here as he preempts the move from multiple contradictory meanings to multiple contradictory versions of an existing entity, or what are sometimes called "ontologies:"

> Can we treat the study of "the nature of human knowledge" just as the study of certain ways in which human beings interact, or does it require an ontological foundation? …The first alternative leads to a pragmatic view of truth and a therapeutic approach to ontology (in which philosophy can straighten out pointless quarrels between common sense and science, but not contribute any arguments of its own for the existence or inexistence of something). … The second alternative leads to "ontological" explanations of the relations between minds and meanings, minds and immediate data of awareness, universals and particulars, thought and language, consciousness and brains, and so on… Such

> explanations, when ontological, usually take the form of a redescription of the object of knowledge so as to "bridge the gap" between it and the knowing subject. To choose between these approaches is to choose between truth as "what it is good for us to believe" and truth as "contact with reality."

The pragmatist approach advocated for here, called epistemological behaviorism, has the benefit of eliminating paradoxical ontological multiples, but sacrifices the ability to put different things, that emerge in different contexts, and are measured in different ways, onto the same ontological footing. Instead, following Wittgenstein, descriptions must attend to the practical involvements at play in the world of the objects described (1968, §193-195). Beyond the therapeutic effect of eliminating paradoxes of what does and does not exist, this pragmatic approach leads discussion toward questions of power and away from questions about where the truth of "self," "agency," "health," or "mood" really lies (cf. Suchman 2011 as quoted in the introduction to this dissertation, for another example of how this idea has been taken up in recent ethnographic work in STS).

While Rorty's broader "epistemological behaviorism" approach is inappropriate here due to its emphasis on social justification, which leaves out the important role of technical and material parts of sociotechnical systems, its critique of representation is otherwise compatible with ethnographic approaches and provides a prescient warning against the problems of representation being reproduced in the realm of ontology. Philosophy, in Rorty's proposition for it, seems similar to the communication studies and sociology traditions in which this dissertation is based—it aims to empirically study interaction to explain discrepancies between common sense and science and between scientists' own theories of what they do and what they are seen to do in observation.

5. Blind spots of the multi-layered participant theory revealed in observation

This chapter has so far focused largely on the formal accounts that emerge as lab members describe their work and present it to peers and colleagues in the field. These accounts are fundamentally limited. While they cohere will larger discourses on health and medicine, and constitute a development on theoretical structures from modern psychiatry, these accounts do not accommodate the situated practices of the lab that constitute their work as it is observed. My observations attend to the messy, contingent ways in which the TPL trial is enacted by particular people in interaction. Some of the of people, data and processes that constitute much of the trial as it is observed are conspicuously absent from formal accounts. The below descriptions of observed activities not only provide a rich, situated account of what happens at the TPL trial (or reference these accounts as established in other chapters of this dissertation), but also remind us of the kinds of activity that are especially difficult to describe using common theories of health, motivation and behavior.

5.1 The missing role of the PWP guide: practices of presence and engagement

The theory of the multi-layered participant, in its emphasis of that participant's underlying agential capacities, lacks a nuanced account of the extent of PWP guidance work. As discussed in chapter 2, PWP guides go beyond technical support, offering a complex form of care work. They explain ML outputs to participants, instruct them on the plan that has been assigned on the basis of these outputs, and then help them to carry out this plan by meeting each week to discuss progress, challenges and goals for next week. This includes, for example, attentive listening as participants share their responses to worksheets about gratitude and interpersonal relationships, which are part of the positivity plan. At the end of six weeks guiding a participant through the exercise plan, one guide asks the participant about social supports in their community and suggests alternative options that could take the place of their weekly conversations.

As is established above in Section 3, descriptions of PWP guidance work are minimized in comparison to the potential for participants to achieve their own goals and address their own issues. The PI suggests that guides are not really the ones "doing the work," as the plan hinges on the participant already having, at the start of the trial, an underlying motivation to change behavior. Journaling is suggested as a way for participants to read what they are thinking, and therefore better understand their deeper self-knowledge. These hidden layers of the participant take the place of robust accounts of guidance work, which includes helping participants to enact recommendations in their own neighborhoods and careers and family arrangements, and explaining ML outputs to participants (for specific accounts of this guidance work see chapters 2, 4 and 5).

Underspecifying the new forms of human labor that are required to implement technology-focused initiatives is a common trend, especially in areas when ML is leveraged toward care work (Gray & Suri, 2019; Oudshoorn, 2008; Semel, 2021). Rigorous accounts of new, non-traditional forms of work involved in digital health programs have been called for in the digital health literature (Koutsouleris et al., 2022, p. 836) and also in Strategy 4.3.B.3 from the latest strategic plan from the U.S. National Institute of Mental Health (*National Institute for Mental Health Strategic Plan for Research*, 2022, p. 45). Drawing attention to these care practices is one of the strengths of ethnographic approaches which combines study of work practices themselves as well as the discourses in which they are conceptualized (for related examples of this approach see Berkhout & Zaheer, 2021; Brown & Michael, 2003).

5.2 Participant-in-the-world: the missing role of environmental factors

In an increasingly pressing problem described at the outset of this dissertation, "most machine learning solutions in mental health are being developed in silos, without the inclusion of decision makers, experts, and end users" despite the fact that "studies have repeatedly shown how

individual technology-based mental health interventions do not exist in a vacuum; rather, they form a part of socially situated and structurally influenced pathways to care" (Koutsouleris et al., 2022, p. 835). The ways in which the force of the TPL intervention is attributed to the multiple layers of the participant serves to isolate the two emphasized components of the intervention, the ML system and the participant, from the material circumstances of the world in which they are enacted. In the first chapter of their edited volume on care practices, Annemarie Mol and her coauthors explain how this separation serves promissory discourses of rational, data-driven technologies: "Care overflowed and was impossible to calculate, technology was effective and efficient. Care was a gift, technology made interventions. Much of the resistance to squeezing care into technological frameworks is informed by this line of thought ... *A life world here, and a system over there.*" (Mol et al., 2010b, p. 14; empahsis mine). While the PWP guidance process served as an entry point for care practices including active listening, it also opened up opportunity for participants to reflect on the contingent, material worlds in which the invention was enacted. In one example, explored more in Chapter 5, a participant who was assigned the exercise plan uses her PWP guidance sessions to explain to her guide things about her experience that he could not learn from the wearable data. The walks she takes in her neighborhood appear in the automatically collected data as steps with a particular quantity and speed. In the self-report data they are some number of minutes of "mild exercise." In a PWP guidance session, though, we hear about how walking around the neighborhood is not the same as walking on a trail, as she finds herself always running into the same few people, leading to sometimes pleasant and sometimes tiresome exchanges. One of the intersections on this route is unsafe, she also mentions. A neighbor was hit by a car there recently. The time of year becomes relevant and she implements the plan in hot summer conditions, though this isn't accounted for in the data system.

Another form of mild exercise, gardening, opens up in her descriptions. She describes its value as based in more than physical exercise: gardening for her is a "mindless" form of "play." Similarly, she notices that observing her grandson's soccer games might not count as "moderate exercise," as she is not the one engaging in physical activity, but being part of a community that centers around this kind of activity is nonetheless a "mood booster."

The contingent, material circumstances of life in the world do not appear in laboratory descriptions of the intervention, yet in this case, the participant explains to her guide the relevance of these topics. The social fabric of the neighborhood converges with the infrastructural problems with streets and sidewalks, and the purposeful yet playful gardening work is assigned a special status beyond its official designation as "mild exercise." The personalized nature of the trial and the theory of the participant explained above supposedly constitute a built-in solution to these variations in environmental factors: "we're trying to encourage the participant to come up with a solution themselves, ideally…they're coming up with a solution that works for them, that's relevant for them and their family situation, with their lifestyle, with their ability to get fruits and grains and vegetables and things like that." The theory of an empowered participant suggests a limited need for taking these factors into account as part of the psychiatric intervention, as participants are free to make their own accommodations as a result of the personalized, flexible nature of the plans.

The PWP guidance sessions offer an opening for discussion of these factors. However, the notion that the participant who attends PWP guidance meetings is just the top layer, with the intrinsically motivated self concealed below, minimizes the value of this talk, which can be attributed to the surface-level version of the participant. In this way, the positing of multiple layers of the participant furthers the issue, commonly noted in sociology of medicine, of the

silencing of the patient by relegation to the realm of a medical object. Mark Sullivan notes that, on the autopsy table, a patient's self-knowledge is subordinated to the supposedly "objective" evidence found in the flesh of the body (1986). In this case, a theory of self and self-knowledge itself is what enables the issue. Also, failing to take environmental factors into account has immediate implementational consequences for participants, as Birk and Samuel have pointed out, describing the assumed "ideal user" of digital mental health technologies: "asumptions can be problematic… One's neighbourhood may not be safe to traverse, or there may be a lack of public transport which decreases one's ability to travel to new places. Or what about homeless people – a group whose high prevalence of mental health problems is well-documented? This is likely to be a group that would fall outside the 'ideal' user" (2020, p. 6; Kahlert, 2016).

Observational studies of the implementation of digital health systems have the potential to reveal the gaps and problems that complicate the deployment of supposedly neutral and portable technologies. While the present work attends to the particulars of laboratory interactions, demonstrating the practical accomplishment of PWP guidance work, for instance, it could go even further in accounted for the situated experiences of participants. The "walkthrough method" for example, produces qualitative descriptions of the experience of using a mobile application in the streets, parks and neighborhoods in which people live and work. This kind of approach, which requires observation of the participants as they walk, run, eat and interact throughout their days, could even better reveal the environmental contingencies of implementing digital mental health programs.

6. Conclusions: theorizing and resolving multiples through observation

In laboratory descriptions of the TPL trial, multiple layers of the participant take the place of other actors, infrastructures, data and processes whose role becomes more clear through observations of lab activities. Ethnographic evidence can also lead the other direction, however.

Annemarie Mol's *The Body Multiple*, for instance, shows how what seem singular in formal accounts are, when seen in practical context, actually multiple. Mol shows how contingent, material circumstances enact diseases, institutions and places: "since the object of manipulation tends to differ from one practice to another, reality multiplies. The body, the patient, the disease, the doctor, the technician, the technology: all of these are more than one" (2002, p. 5). Two different diagnostic procedures, carried out in different circumstances, can multiply the ontology of the disease (2002, pp. 141–142). Observation of practices at the TPL trial shows how PWP guides are relevant actors, and participants themselves do not contain all of the roles they are said to in formal accounts. In Mol's case, attending to concrete practices leads her to suggest multiples rather than challenge them.

Following Mol's attention to heterogenous practices, but working in the domain of critical data studies, Ericson & Haggerty aim to account for the heterogenous forces or "flows" that constitute what they call the surveillant assemblage, extracting data about bodies and activities and reconstituting them into "data doubles" (2000). Describing the forms of alienation emerging from the big data economy, Mark Andrejevic describes the data collected about an individual as their "data shadow," a result of their labor from which they are alienated to the point of incomprehension (2013). Minna Ruckenstein builds on these ideas, critically exploring how people conceptualize and reflect on their data doubles, which, in their various visual and graphical manifestations, take on their own agency as people rely on them for self-understanding (2014a). These theories, like Mol's, differ from the formal accounts at the lab in that they, by design, exceed formal accounts to question the practical, contingent "doing" of data science initiatives, yielding theories of "doubles" on this basis.

The present chapter contributes to this literature and, more broadly, what Kristensen and Ruckenstein call "the anthropological tradition that challenges notions of the individual and the self as bounded entities with a fixed and stable ontology" (2018, p. 3627; see also: Nafus, 2016). Lab members of the TPL describe their intervention as primarily carried out by an empowered participant who contains hidden layers or versions of their self which can be activated by the insights of the ML system. This formal account, which differs from the intervention as it appears in my observations, is part of a development in psychiatric theory whereby the notion of multiple, often contradictory meanings of the psyche or the self is replaced by multiple selves or ontologies. In Richard Rorty's pragmatist critique of representational epistemology, he points out this ontological doubling as a new version of an old philosophical attitude toward knowledge as only accessible through more or less veridical representations of reality.

Finally, the multi-layered theory of the participant is contrasted with my own observations, demonstrating the importance of aspects of the digital mental health invention like the novel, underspecified work of PWP guides, and the environmental contingencies that constitute the experience of implementing a PWP plan in the real world. Further, these observations, put in contrast with formal accounts, demonstrate the kinds of knowledge about the experiences and interactions of participants and PWP guides that are not accommodated by the prevailing theoretical explanation.

Acknowledgements

Chapter 3, in part, has been submitted for publication of the material as it may appear in *Social Theory and Health*. Berners-Lee, B. The dissertation author was the sole researcher and author of this paper.

Chapter 4 The Social Organization of Guidance Training

1. Introduction

Having established the important, if overlooked role of the PWP guide, this chapter will attend to the particular interactions in which they are trained for guidance; the following chapter, Chapter 5, will show how this training is put to work in their interactions with participants.

As PWP guides prepare to offer guidance to participants, they meet with one or both PIs to learn more about this process and ask questions. This chapter investigates the kind of interactions involved in this process, using an observation approach to understand how the PIs of the TPL interact with guides during training. PWP guidance meetings feature interactions that are largely limited to the topic of managing participant resistance and equipping the guide with knowledge of what to say in response to queries from participants. The role of the guide is curtailed by characterizing PWP guidance as primarily constituted in the task of saying the right thing to the participant—a communicative activity isolated from the practical context of the intervention and participants' attempts to implement it in their lives.

Chapter 3 of this dissertation describes how, in formal accounts of the trial, environmental factors are minimized and the intrinsic capacities of the participant are multiplied. In the present chapter we see how the PIs theory of PWP guidance, as it is expressed in trainings, similarly skirts discussion of how participants' living and working environments could affect the trial. As ethnomethodological theories of interaction suggest, the practical circumstances about

which interlocutors interact inevitably influence and help constitute the course of that interaction—conversational strategies and topics of conversation are not separable in practice (Garfinkel, 1967). In PWP guidance trainings, however, interaction with participants is made to seem like a question of *saying the right thing*, without any need to address or account for the practical circumstances that make up the context of guidance sessions. This finding also coheres with chapter 2, which demonstrates how guides are made to seem like a component of the technical apparatus of the trial in formal accounts, and investigates how this account of their work affects the way they talk to participants. These two aspects, the isolation of guidance communication from the broader circumstances of the trial on one hand, and the minimization of guidance labor as technical support, are linked below using the theory of legitimate peripheral participation as expounded in Lave and Wenger's influential book (1991).

The excerpts of interaction provided in this chapter make clear that guides are instructed on how to respond to participant questions and complaints, even when they raise questions about the design of the intervention and the protocol for implementing PWPs. The below analysis shows how this is accomplished in conversation, attending to the specifics of interactions as they were observed. Hypothetical active voicing (HAV), a kind of reported or quoted speech that refers to a hypothetical future scenario, is shown to be a particularly relevant conversational resource for relegating the discussion to the issue of how to best respond to participant queries. The below analysis shows the relevance of HAV and other conversational resources, and also how they are marked linguistically. In this way, the present chapter addresses the question of how the PWP guidance role is established in interaction.

2. Literature review: the interactions of training for a future scenario

Relevant existing research in this area concerns planning and training activities, which are by nature oriented around producing competencies to be deployed in future scenarios. Studies of

instruction and training that use conversation analysis methods tend to have their theoretical foundation in theories that see learning as the result of participation in social processes, including entering or establishing a community of practice (Lave & Wenger, 1991). While some studies that use this participation framework focus on cultural meanings and identities (Lave, 1997), those taking an observational approach often emphasize the situations in which, and for which, a trainee is taught, seeing concepts and knowledge as manifested in collaborative activities that rely on aspects of the situation including environments, artifacts, gestures, bodies and forms of multimodal interaction that emerge from a dynamic interplay of these factors (Goodwin, 1994; Koskela & Palukka, 2011). Leyland's study of lesson planning, for example, follows Lucy Suchman's influential theory of plans and situation actions, according to which plans are not representations of actions to come, but rather creative discourses that are inevitably incomplete and contingent with respect to the actions they describe (2007). Leyland investigates lesson planning using a conversation analytic approach to specify how, through talk, gesture and manipulation of objects, teachers orient themselves toward and embed themselves within the future teaching scenario for which they are planning (Leyland, 2016). These literatures show the complexity of instruction and training which follows from the fact that the tasks for which trainees prepare cannot be easily isolated from the contingent doing of the practical task. The present study exemplifies an attempt to reduce this complexity by presenting the task of PWP guidance as simply responding appropriately to participant inquiries.

Other existing literatures develop theories on the interactional role of imagined, projected or embedded scenarios, laying groundwork for our discussion of conversational resources for PWP training, including HAV. Speakers use particular interactional resources to refer to, report on, and enact talk from past experiences, fictional or imagined characters or imagined future

scenarios. This kind of quotation of talk that enacts interactions other than the present conversation has been referred to as direct reported speech, or DRS (Holt, 1996). As will become clear in the below analysis, however, this kind of speech does not so much report or forward some other talk, but rather enacts a past, future or imagined interactional context, affecting what Goffman (1981) calls the "footing" of co-present interlocutors. The enactive aspect of quoted talk is clear when it shifts the conversation to a hypothetical future scenario. Referring to, enacting or quoting imagined future interactions was termed by Simmons and LeCouteur "hypothetical active voicing" (HAV), drawing attention to the role of the speaker in giving voice to the figures of an imagined future interaction (2011).

The present study takes interest in HAV as enacting a future scenario. Specifically, uses of HAV in the excerpts below take place in a PWP guidance training meeting, and serve to enact the PWP guidance sessions themselves for which the guides are preparing. Among various reported uses of HAV, it has been described as sometimes taking the form of hypotheticals that ambiguate whom is expected to implement advice (Silverman et al., 2004). Instead of seeing HAV as a way of ambiguating the interactional context of advice giving, however, the present study takes interest in how it enacts the future guidance scenario, and thus shifts the footing and interactional context, embedding quotations from an imagined future guidance session into the training meeting. To use terms from Goffman (1981), the PI "animates" an imagined future version of the "figure" of the guide. Recent research that develops Goffman's theory of this conversational resource includes Zuckerman's (2021) study on what elements of represented discourse appear when animators embed figures in their speech, and, building on a concept from Agha (2005), how "transparently" these figures are heard as such. The process of embedding past, fictional or imagined future talk into interaction also has been shown as a gestural

accomplishment (Leyland, 2016) and one that makes use of spatial features of the environment (Goodwin, 2003). Other recent research demonstrates how this kind of talk makes use of artifacts which are selectively transformed as speakers change footing. Of particular relevance to our case, however, are not "object-focused" discourses through which, for example architects discuss and coordinate their talk around an imagined future building (Murphy 2011), but rather discourses that focus on thought, affect, emotion and "inner speech," as these constructs are particularly relevant to the digital psychiatry intervention.

Research on the use of quotation and HAV in the psychiatric context includes Barnes and Moss' study of voicing past inner thoughts, which includes excerpts from psychoanalytic psychotherapy sessions (2007). Particularly relevant to the present study, Simmons and LeCoureur (2011) discuss HAV in the context of cognitive behavioral therapy (CBT) sessions in which therapists use the conversational resource to model healthy responses and self-talk (Leahy, 2001). The authors explain some of the ways in which HAV is marked and how it serves to pre-empt resistant responses from clients along the lines of "I wouldn't know what to say." Because the therapist has modeled the recommended conversational strategy, clients are unable to resist on these grounds. Client resistance, however, still frequently occurs in their study, but it is not based on a lack of clarity on how the patients should talk with themselves or others.

This work is similar to the present study, as it shows how therapists use HAV to ensure that their patients always have a sense that they know what to say, either to their self or to others. While these authors suggest that demonstrating what patient should say in the future pre-empts resistant responses, our study instead suggests that HAV can be used to center the discussion around questions of what to say and when to say it, not preempting the response " I don't know what I would say," but rather preempting questions about how practical and logistical factors

might complicate the practice of PWP guidance—factors which make the interactional challenge of PWP, in practice, more than just saying the right thing to assuage participant questions and complaints.

3. Methods

The below data were collected as part of the broader observational study of the development and testing of a machine learning based digital phenotyping intervention for mental health. The present analysis focuses on meetings in which guides were trained by the PIs of the laboratory. These meetings were conducted on the Zoom video conference platform and in one case, in a hybrid format in which several laboratory members were co-present in a meeting room that was shown on video conference for remote attendees. Meetings were recorded using screen recording and then transcribed using ELAN qualitative data analysis software.

I took field notes on observed interactions and used tags to indicate emerging themes and reveal patterns that could constitute relevant conversational resources or interactional repertoires. Excerpted interactions were selected using these tags and then transcribed using conventions of the field of conversation analysis (Jefferson, 2004). HAV is analyzed here as an interactional accomplishment that emerges from the dynamics of training sessions, rather than an *a priori* linguistic structure (Goodwin, 1979). Rather than take a conversational sequence or linguistic construct as a "topic" to be investigated at the field site, these linguistic resources emerge as answers to questions about how the common-sense doing of the training takes place which motivate the study (Garfinkel & Wieder, 1992).

These methods were selected for their suitability in addressing questions around how the social organization of PWP guidance is accomplished. While it is immediately clear that PWP guidance training includes several moments in which guides' concerns about implementing the trial are met with a response that centers around how to talk to participants, close analysis of

these interactions is required to demonstrate how this shift is practically carried out and the implications it has for the role of PWP guide.

4. Analysis

To understand how the role of the guide is established in training, we first look at instances in which the guide raises questions and issues that relate to the implementation of the trial broadly speaking. In these cases, PIs respond by shifting the conversation toward the issue of how to respond to participant questions and complaints. Having demonstrated this pattern, I go on to show how this is accomplished through conversational resources including HAV. Finally, I will show how HAV is marked using linguistic particles, and the role these markings have in shaping the use of HAV and the social organization of PWP guidance trainings in turn.

4.1 Conversational resources to establish the PWP guidance role as responding appropriately to participant queries

Focusing on the parts of the below excerpts in which guides ask questions about the broader implementation of the trial, we see how particular conversational resources work to shift the topic of conversation away from study design, trial protocol and other research implementation issues and toward discussion of how to resolve participant complaints as such.

In Excerpt 1, the guide asks about cases in which participants have disabilities, which he calls "A:ctual problems," on line 3. He goes on to ask if this kind of case might require reassignment to another plan — if participants with disabilities should be assigned the second-ranked plan (as determined, in part, by the ML system), in cases in which the disability could interfere with the participants' efforts to implement the plan (lines 6-7).

The PI responds by reiterating an idea about participant motivation that he has shared before. As discussed in chapters 2 and 3, the PI considers intrinsic motivation to be a kind of safeguard against imposition of inappropriate recommendations. Earlier in this dissertation the issue was discussed as part of the larger formulation of the multi-layered participant, who forms one side of an intervention characterized as consisting primarily of individual participants and

the technology that they wear and that is used to analyze their data. Here we see one way in which this outlook is constituted in interaction during training.

Rather than expound the theory of intrinsic motivation, the PI shifts to HAV, first, on lines 12-13, warning against reassignment using HAV, “I don’t think it ever would make sense that we would say…”, and then, on line 18, switching into HAV again to demonstrate a preferred response, “the first session is always gonna be that discussion of…”. Later, before the guide has had a turn, the PI continues to issue another recommendation using HAV, this time explicitly labeling it as a kind of quotation, saying “the ideal quote is” (line 40).

Excerpt 1.Questions of disability during PWP Guidance training.

G: I had a question about the exercise program if we're gonna move uh uh from it. Ehh. One is that well some of these patients are gonna have disabilities. A:ctual problems that eh is gonna keep them from exercising or (.) doing that 20 minute cardio.

PI: Um hum=

G: =E:h. In these situations do we like eh suggest them the ↑second suggested plan for eh second best suggested plan um eh that we were talking about? E::h. Do you have any suggestions about that part of it?

PI: That's a great question. So y'know the the way I would <conceive> of this is that (.3) um at some level a person has to have some inherent desire or motivation↓ to change the behavior that's being targeted here. So (.3) I don't think it ever would make sense that we would say ↑hey you should be in a plan (.) or this plan unless they themselves were buying into it and show choosing or electing that plan too.

G: (Ma::y b: tha)=

PI: =uhm. So that's kind of the the first point is like the the plan assignment and I think that is really the first session is always gonna be that discussion of ↑hey based on your data this is the number one thing that seems to relate to your mood. Y'know and what do you think. And kind of just seeing if somebody has some interest in (.) selecting a plan to target that behavior or not. Uh (.7) and if they don't, absolutely go to the ↑hey well then the second thing that could be recommended based on our model that is >associated with mood is< and then whatever the number two thing would be, right. So I think that's like the first session is (.) uh having some buy in and having some motivation that they're like yes this is a particular aspect of my behavior I want to change. (.) Um. I think the second part of your question is that <at some level whatever they've entered into the data> that goes into our model suggests they can do (.) that. Right so sort of where we're trying to at least for the activity part of where the personalization comes in (.) general general clinical trials in this ↑area might have something on the order of like ↑we recommend everybody to get 30 minutes a day of moderate intensity exercise. Or (.) y'know we'll do a trial where we look at ↑weight lifting or specific kinds of ↓exercise.(.4) And here what we would be doing is actually tailoring

```
36        our recommendation with a target in mind. The target might be on
37        average 150 minutes per week of exercise.
38  G:    Um hum=
39  PI:   =but really starting with wherever they are. Right so y'know that's
40        where like if you look at the exercise um (.) the the ideal quote is
41        your data suggests that y'know X exercise for X duration at X intensity
42        for this many minutes per day is good for you. So that that data will
43        really be based on their own their own reports and and their own
44        monitoring of their data. And it could be something on the order of (.)
45        y'know we've reviewed your data and we find that whenever yo::u walk
46        ↑briskly for 20 minutes your mood is is improved after that (.) or 15
47        minutes.
48  G:    [Um hum]
49  PI:   [O::r  ] uh having:g this many steps per day is associated with an
50        improved mood. So the the personalization is kind of like starting from
51        where they are (.) but with an end goal of trying to get somebody in
52        the habit of exercising y'know 20 to 30 minutes per day.
```

The PI continues by suggesting what to say if the participant doesn't have interest in the recommended plan (lines 22-24). In doing so, the PI suggests that it is acceptable to offer the secondary recommendation (as listed on ML outputs). The PI uses HAV to suggest that this switch to the second choice could be an outcome of the conversation in which a participant offers resistance. The PI describes this view later in the conversation, on line 38, as "starting wherever they [the participant] are."

In these moments of talk, the PI further establishes the purpose of the conversation as providing the guide with strategies for communicating with participants and coping with participant resistance as necessary. The PI agrees that a secondary plan could be assigned, but does so in the context of managing participant resistance and maximizing the participant's intrinsic motivation (lines 10-11). A reassignment could be justified not because of the direct clinical relevance of disabilities as they interact with mental health, but rather because of how the

relationship between disability and mental health might affect the motivation of the participant. The PI needn't fully explain the theory of the participant discussed in chapter 3, though, as he uses HAV to orient the discussion around managing participant questions, rather than the design of the intervention and its accessibility to people with disabilities. The formal account of personalization and individual motivation is paired with an interactional repertoire through which guides are trained not to raise issues about how external factors could stymy participant attempts at implementing the PWP, but rather communicate ML outputs so as to elicit participants' inner motivation, self-knowledge and positive mood.

Excerpt 2. Question of nutrition counseling during PWP guidance.

G: Could there ever b:e ((clears throat)) a situation you think (.) where somebody has (.) a specific question and wants nutrition counseling are we in a position to (.) (like) guide the:m (.) accordingly and in those circumstances where they want (.) professional opinions?

PI: that's a pretty good question um. I I would probably say not not necessa:rily or not really. What would be the context of that? Like if somebody said I'm not sure if I'm getting enough protein or something? like what would be the context in which somebody would want a nutritiona::l (.) expert here (.) that you were thinking=

G: =maybe yeah I don't know if somebody is skeptica:l about certain changes and (.) once (.) I get somebody that has a particular nutrition >background for< either the changes that they're (.) planning on undertaking or (.) maybe even midway through the the intervention and they feel like they want (.) ↑I I don't think it'll happen necessarily but I was just [cu: curious to see if]

PI: [Um hum. Um hum]

G: Like we would eve:r like consult anybody for a nutrition specific advice.

(1.8)

PI: That's not how we've, we conceptualized this par[ticular trial] idea.

G: [Right. Yeah.]

PI: Um. (1.1) I think it could I mean nutrition is one of these topics. I think at some level all of these topics but nutrition in particular >↑pe[ople have<] very strong [opinions on] y'know[people have]

G: [Right] [(h)Yeah(h)] [(clears throat)]

PI: you'll have like y'know the protein group or the Atkins group or the Paleo group or the you kno:w (.) low fat group and each of them feel like (.) there's their truth is the only truth and y'know the the the way the health is (.6) that particular approach to eating (.4) Um. You'll have people who are y'know now it's like the chronobiologists or you'll be like only eat within an ↑8 hour segment of time (.9) I I think (.6) if people raise questions and say like well ↑I don't think this is the right ↑nutritional plan for me I don't know is the right diet plan.(.3) I I ↑wouldn't necessarily say our role is to <push back>. Y'know ultimately I think our role is to say (.5) ↑here is one (.) nutritional recommendation that's been tested for mood↓. So we're

not trying to make any:y claims that (.) this particular nutritional plan is going to b::e (.8) good for ↑weight loss or ↑energy or ↑sleep or any of the other things that some (.) some other people might be cl↑aiming. We're, We're just saying ↑hey look (.) uh or the ff the approach °I would take is like (.) y'know° this is something that's been helpful that's been shown to be helpful for mood (.2) um there are reasons why this particular:r diet is helpful for ↑mo:od and y'know may relate to anti-in↑flammatory benefits o:f (.) whole grain and vegetable based and other things uhm (.) but really at the end (.) y'know you can say if it's not working what parts of it are not ↑working what parts of it do you think ↑are working and just try to kind of work within that (.) that's how I would approach (.) any >skepticism or pushback< is to >try to kind of work within< (.) the context of >what what do what< do ↑people find (.4) and y'know ↑ultimately (.) part of (.) the rationale for all of these >and ↑even in< y'know the example ((name of other PI)) made (.) y'know one of the specific things they look at are the::y >they're asking information about is ca↑ffeine intake< (.1) so there may be some specific things that don't aren't part of the ↑general recommendation (.) but still might come up in the model as specific recommendations that (0.8) we might give to participants.

In Excerpt 2, the guide starts by asking a question in the context of a future interaction with a participant. The guide asks the PI about a situation in which a participant asks for nutrition counseling. As he further explains his question, the guide asks if the lab could consult a nutritionist to provide specific insights on food and diet (line 17). As was the case in Excerpt 1, the guide here raises a question regarding external factors that could influence or alter the course

of the study, rather than staying within the topical bounds of how to speak with participants during guidance meetings. The PI responds first by explaining why food and diet might be controversial topics with participants. The PI then uses HAV to suggest what a participant who is resistant to dietary recommendations might say and to model an appropriate response as a guide (lines 33-44).

As we saw in the first excerpt, the PI uses HAV to shift the topic away from the trial protocol, which does not include consultations with experts in nutrition, reorienting the conversation toward the challenge of responding to a participant. Along with explanations of the theory of the participant as empowered by the ML technology (explored more in chapters 2 and 3 of this dissertation), PIs also use conversational strategies to shift discussion away from the question of whether implementation of ML outputs for specific individuals might require the additional labor of specialized clinicians (for discussion of how sociotechnical systems that are described as ML or automation based can sometimes require more human labor, not less, see Chapter 2).

In excerpt 2 we see how HAV combines with other conversational resources to help the PI make an argument in favor of the particular dietary recommendations of the PWP plan as opposed to the other nutritional plans and theories he describes on lines 25-29. The PI's use of HAV to describe the benefits of the plan, on lines 39-42, serves to distance the PI's endorsement of the nutritional plan from the guide's questioning of it. By expressing the nutritional value of the plan, including "anti-inflammatory benefits," in HAV, the PI directs the response to an imagined future participant, allowing the guide to save face in case they are unaware of these benefits, as the PI is not only sharing the nutritional benefits of the plan with the guide, but also modeling how to explain them to a future participant. The use of HAV is paired with the vague

warrant "it has been shown," allowing the PI to avoid discussing the clinical complexities of nutrition with the guide and to instead model the vague, authoritative discourse of a guide conversing with a participant[15]. By responding to a question about nutrition as if it were a question from a participant, the PI avoids deeper discussion of nutrition as they are modeling appropriate talk for guidance meetings, rather than discussing the intervention itself with a colleague.

4.2 Marking HAV in guidance trainings

The way that HAV reorients discussion around the task of responding to participants emerges partly from the way that it is marked—the ways PIs signal this shift in what Goffman calls "footing," from speaking directly to a guide, to speaking in the role of a guide to an imagined participant in an imagined future scenario. By investigating how HAV is marked, we get a better sense of how it is possible to shift conversation from a direct exchange to a demonstration of some suggested future talk, and we see how this future scenario is embedded into the primary footing of direct exchanges between PI and guide.

To shift into HAV, PIs make use of markers including initiating reporting verbs like *say* and *saying* (Excerpt 1, line 13; Excerpt 2, lines 30, 37 and 43), the particle, *y'know* (Excerpt 1, lines 20, 40, 43, Excerpt 2, lines 39, 41, 43) and an upward shift in prosody (Excerpts 1, lines 13 and 22). All of these features were noted by Simmons and LeCouteur (2011, pp. 3179–3180) as strategies for marking HAV in therapist–patient CBT sessions.

4.2.1 *Hey*

Another common marker of the switch into HAV is the use of *hey* to indicate that HAV has started. This kind of marker is not reported elsewhere in the CBT literature. Following the

[15] Simmons and LeCounteur report a similar function of HAV, as marked by *k'now,* serving to frame information as common knowledge and thus preempt resistance from patients (2011, p. 3181).

common-sense meaning of *hey* as a greeting and vocative, in this case it serves to open the future speech enacted by HAV. While *hey* constitutes reported speech or active voicing as it suggests an imaginary hearer other than present interlocutors, it does not serve an initiating function in the imagined future scenario, but rather indicates to the co-present guide that HAV has been initiated.

Seen through the schema from Goffman's "footing," in which "animators" give sound to the speech of other "figures" that they represent, we might think the interjection *hey* serves as a greeting—the PI animates the figure of the future guide as greeting the participant. The intended recipient of the interjection *hey*, however, is just as much the co-present guide, to whom it indicates a switch to HAV. The *hey* marker does not prepare the co-present guide in advance for a switch into speech which will soon no longer be directed to them, but rather simultaneously serves as both content of HAV and a marker of the change in voicing that has just occurred. If this were not the case, HAV speech would seemingly occur only at the very beginning of a guides' interactions with participants, at which point a greeting is appropriate.

Hey, as used here as a marker that the PI has shifted into HAV, serves as a bridge between the present guidance training and the future guidance session. In his chapter on footing, Goffman discusses how the *I* in statements like "I once lived that sort of life" simultaneously serves multiple footings, "bridging between the scene in which the talking occurs and the scene about which there is talking" (1981, p. 150). Goffman goes on to suggest that this "bridging power" is not unique to *I* but also applies to other pronouns. In this case we see how *hey* similarly serves to simultaneously mark initiation of HAV for the co-present hearer, and as vocative of the imagined participant (Ameka, 1992, p. 115)—thus the initiating and vocative functions of the particle are used together to signal a change in footing, and to address an

imagined participant, with the former function operating in the footing of the training and the latter in the imagined guidance scenario that is embedded within it.

4.2.2 *Y'know*

Following the literal meaning of the expression "you know," *y'know* has been explained as serving to establish the "information state" of a conversation, helping the speaker determine which parts of a discourse are shared knowledge, and which parts might be new to the hearer. In our case, however, *y'know* serves as what Schiffrin calls an interactional marker: "Because *y'know* induces a hearer to act as an information recipient, it has the complementary function of ratifying the speaker as an information provider but one in a somewhat specialized production format - one whose provision of information is contingent upon hearer reception" (1987, p. 274). Rather than a legitimate prompt for the hearer to share their familiarity with a topic, *y'know* here serves to remind the hearer of the information-giving footing of the speaker.

Michael Silverstein's concept of orders of indexicality (2003) can help "relate the micro-social to the macro-social frames of analysis" by showing how indexes on any given level "n" can also index, when considered on the level n+1, something about what Silverstein calls the appropriateness of the utterance having been said (to a particular hearer, by a particular speaker in a given context). In the case of *y'know,* if we consider information-giving itself the first order, then the establishing-status-as-information-giver is the second order indexical function. It prepares guides to hear a piece of discourse that is suggested as a model for guidance, and it is interspersed into HAV to remind the hearer of this footing. In this way, it indexes not only the speaker's status as information-giver, but also the hearer's status as information-receiver (Goffman, 1981, p. 151). Thus, as in the examples from Silverstein, the second-order index is at odds with the lower order literal meaning of "do you know" because of how it can forestall the hearer's next turn rather than invite it.

Just as there may be overlapping particles like *hey* that seem to play distinct roles in each of the footings they simultaneously occupy, the case of *y'know* also demonstrates how footings are overlapping and intertwined. There is a continual "holding at abeyance" (Goffman, 1981, p. 155) of the training footing of the PI, while HAV is used to temporarily evoke the embedded future guidance session:

```
37  PI:  some (.) some other people might be cl↑aiming. We're, We're just
38       saying ↑hey look (.) uh or the ff the approach °I would take is like
39       (.) y'know° this is something that's been helpful that's been shown to
40       be helpful for mood (.2) um there are reasons why this particular:r
41       diet is helpful for ↑mo:od and y'know may relate to anti-in↑flammatory
42       benefits o:f (.) whole grain and vegetable based and other things uhm
43       (.) but really at the end (.) y'know you can say if it's not working
44       what parts of it are not ↑working what parts of it do you think ↑are
45       working and just try to kind of work within that (.)
```

In these lines from Excerpt 2, the PI opens HAV on line 38 with a pause and *y'know*, and then offers an example of what to say to a participant. On line 43, the PI pauses and then switches out of HAV, saying "but really at the end," only to switch back into HAV using the reporting verb "you can say," and another *y'know* (see Excerpt 1, lines 19-21, 31-34 for more examples). This example of HAV demonstrates how *y'know* can be used not only to initiate HAV, but also to continuously remind the hearer of the embedded nature of the HAV footing. *Y'know* serves to switch into a voicing that evokes an imagined future scenario, while also establishing that this footing is embedded within the footing of a trainer providing information for a trainee, thus accomplishing what Goffman calls a layering or lamination of conversation (1981, p. 156).

We can get a better understanding of the role of HAV in guidance trainings generally by following the example of how the *y'know* particle marks HAV by laminating the footing of

direct instruction by the PI to the guide and the footing of an imagined future guidance session. Particles like *y'know* serve not to check an interlocutors level of knowledge in order to more effectively share information (following the literal meaning of the expression *do you know*), but rather as a second-order index of the footing from which such a check might be necessary: the footing of the information-giver-who-checks-for-understanding. Not only the footing of the speaker, but also that of the hearer are indexed, indicating to the hearer that they are playing a role of information recipient in this moment of talk. HAV itself can be similarly considered on two orders. The first order consists of the PI supplying the guide with discourse on which the guide can model their own guidance talk. The second order consists of the establishing of the guidance preparation meeting as a place for rehearsing an imagined future scenario. In this way, the second-order indexical function of the voicing can be understood as a way in which, without the need for explaining an elaborate conceptualization of the participant and technologies, guidance labor is made to seem like a technical task isolated from issues of coping with external problems and contingencies that arise as the intervention is deployed in participants' everyday lives.

5. Conclusions

PWP guidance trainings are not for interactions regarding the implementation of the trial generally, but rather the specific task of preparing guides to respond appropriately to participant queries. This structure of the meeting is maintained through particular conversational resources including hypothetical active voicing, or HAV. The notion of HAV builds on what has been termed direct reported speech (DRS) through which speakers forward talk they've heard in another context. HAV, however, draws attention to the inevitably creative, productive nature of what is sometimes called DRS, as this kind of "quoted talk" is set in an imagined future context. In HAV, the speaker engages in what Goffman calls a change in footing, playing the role of the

guide and addressing an imagined future participant. This change is marked by the use of reporting verbs (eg. "I'd say"), particles like *hey* and *y'know*, and prosodic markers including an increase in pitch. The way HAV is marked demonstrates how single lexical items can form a bridge across or otherwise ambiguate different footings, and also how the initial footing of the PI, which is based on direct advice- giving to the co-present guide, is continually kept at abeyance even during long stretches of HAV. Further, *y'know* is shown to operate not as a genuine check for understanding, but rather as a second-order index that works to establish and renew the participation framework of the meeting by reminding the guides of the PI's role as information-giver and their roles as information-recipients.

HAV itself also plays a role in the participation framework of the guidance meeting and, by limiting guides' participation, it also restricts the topics of the training. Specifically, the PI uses HAV as a response to guides' questions about potential modifications to the trial. In this way, the PI shifts the topic away from study design and trial protocol toward the topic of responding to participants in guidance meetings.

5.1. Legitimate peripheral participation

Better including coaches, guides and patients in the development and deployment of digital health technologies has been suggested as a way to make data-driven medical interventions more robust in diverse real-world settings (Koutsouleris et al., 2022). The above analysis of HAV demonstrates the importance of considering the pragmatics of conversation that constitute participation frameworks, especially when working to understand and improve the role structure of a training environment in which trainees, in this case guides, learn how to participate from an instructor, in this case the PI. The boundaries of staff members' involvement emerge in the specificities of laboratory interactions.

Lave and Wenger's influential monograph on communities of practice and legitimate peripheral participation considers the professional roles of trainees and instructors as they emerge in interaction and professional practice. These authors base their inquiry on the notion that "the ability of a community to reproduce itself through the training process derives not from the doctrine, but from the maintenance of certain modes of coparticipation in which it is embedded" (1991, p. 16). For Lave and Wenger, the knowledge that trainees learn is manifest not in an inner cognitive structure but rather in competencies as demonstrated in practice[16]. Legitimate peripheral participation is a framework for apprenticeship that values both the practical involvement of trainees and their position in the periphery: "peripherality, when it is enabled, suggests an opening, a way of gaining access to sources for understanding through growing involvement" (p. 37).

The present study not only contributes to this research program by specifying the interactional moves involved in training PWP guides in the role, but it also can benefit from Lave and Wenger's perspective, using it to consider the training session as a place of legitimate peripheral participation. This perspective, in some ways, supports the PI's choice to model HAV guidance using HAV, demonstrating the kinds of talk that will be appropriate for the guide to use once they begin their work in this capacity. This kind of demonstration better equips guides to participate compared to a didactic form of instruction that describes the appropriate talk in the abstract (Lave & Wenger, 1991, p. 109).

[16] While Lave and Wenger see knowledge as it manifests in practices and interaction rather than as accumulated in an individual's mind, some common approaches to training combine these perspectives. For example, the notion of combinations of vertical and horizontal expertise, which refer to individual skills and "knowledge that is co-created through interactions and movement across contexts" respectively, has been used to study the important role of coaches in helping teachers to interpret data collected by their school and school district (Marsh et al., 2015).

That said, the use of HAV also serves to continually reorient the training session around the issue of *what to say and how to say it* at the expense of discussion of other factors, like the inclusion of a nutritionist or consideration of disabilities, that, in a broader sense could affect the course of guidance. The issue of *what to say and how to say it* is isolated as a communicative skill apart from the complex factors that explain why a particular plan is recommended. These factors are made to seem external to the question of how to effectively guide the participant. The fact that the guides' questions about these external factors are not entertained in the context of study design or trial protocol shows the limits to their role as legitimate peripheral participants: "Legitimate peripherality is important for developing 'constructively naïve' perspectives or questions. From this point of view, inexperience is an asset to be exploited…Legitimacy of participation is crucial both for this naive involvement to invite reflection on ongoing activity and for the newcomer's occasional contributions to be taken into account" (Lave & Wenger, 1991, p. 117). In our case, the guides' questions about nutritionist support and disabilities do not invite reflection, but are rather transposed into the realm of *what to say and how to say it*, as if the guides' concerns were those a participant raised in a guidance session. Taking Lave and Wenger's approach, this appears as a missed opportunity for the PI to reflect on the trial and a limitation on the guides' legitimacy as a participant in the lab.

The approach of this dissertation is based on the notion that communication manifests in the practical accomplishment of interactions as they occur in context, rather than as portable schemas that can be applied across contexts. In line with this perspective, Lave and Wenger insist that "instructional assistance is not construed as a purely interpersonal phenomenon; rather…learning must be understood with respect to a practice as a whole, with its multiplicity of relations — both within the community and with the world at large" (pg. 114). In chapters 2 and

3 we discussed how formal accounts of the trial fail to include material circumstances and the complexity of the interactional work of guides. Here we see how this manifests in interaction, as the PI instructs guides on *what to say and how to say it* as separable from practical contexts around additional staff and participant disability. In Excerpt 1 we see how the two operate in tandem—the conceptualization of the trial (which, in chapters 2-3, I call a "formal account") is expressed in lines 10-11, before the PI switches into HAV. Relegating of the role of the guides using HAV, and foregoing reflection on "external factors" like disability coheres with the notion (discussed in Chapter 3) of the participant as multi-faceted, intrinsically motivated, and already containing the key components of the trial.

5.2 Managing participant resistance in CBT and through guides

The function of HAV as orienting the conversation around the problem of responding to participant resistance reflects findings from conversation analytic studies of CBT. In the traditional CBT context of one-on-one talk therapy between a therapist and a patient, Simmons and LeCouteur show how HAV is commonly used once a participant starts to demonstrate some resistance. It serves to pre-empt responses like "I don't know what I would say," in these contexts. In our case, the problem of participant resistance is emphasized and the same conversational resource is used to train guides who might have to cope with this form of resistance. While it is used in a different context here, it still has the potential to prevent guides from offering a response to the PI along the lines of "I don't know what I would say." In our case, however, orienting the conversation around *what to say and how to say it* (as an isolated capacity separate from the practical context regarding the plan that is being recommended) is shown to limit the role of the guide and foreclose reflection on factors that are deemed external.

The function that Simmons and LeCouteur attribute to HAV in traditional CBT is the element of the guide's interactional work that is made focal in the training. As HAV is

transposed from therapy sessions to the training of guides, the task of preventing and coping with resistant responses from participants shifts from one of many strategies involved in a complex therapeutic interaction, to the central competency being trained for. The PI explicitly acknowledges this competency as a necessary supplement to an intervention which, as is described in Chapter two, is described in formal accounts as accomplished primarily through the capacities of the individual participant and new technologies.

Chapter 5 Coping with Breakages and Contradictions in Self-Reported and Passively-Sensed Data

1. Introduction

As discussed in earlier chapters, digital phenotyping and other digital psychiatry interventions, including the one being developed at the TPL, are based on correlations between mood and behaviors as they are "passively sensed" by gyroscopes, accelerometers and other sensors, without any input from the user. To produce these analyses, however, the subjective experience of mood and exercise experience also must be measured. Indications of mood, along with other subjective accounts like that of exercise duration and strenuousness, are collected through participant self-reports. At the TPL, participants use wearable devices that passively sense their activities and prompt them for self-reports of their mood and exercise. These self-reports constitute an Ecological Momentary Assessment or EMA, which offers a snapshot of how a participant is feeling at a single moment in time (Figure 1). For daily EMAs, participants are also prompted to share their mood how much mild, moderate and strenuous exercise they have taken part in over the last 4 hours (Figure 2).

The present chapter takes interest in how participants in the digital phenotyping study conceptualize their self-reports of mood and exercise, and how laboratory members cope as these self-reports come into conflict with passively sensed data or their own intuition. The study is based on observations of two settings: guidance sessions in which laboratory members support participants as they attempt to implement a behavioral modification plan (PWP), and the

Figure 1. Self-report screen for mood as participants see it.

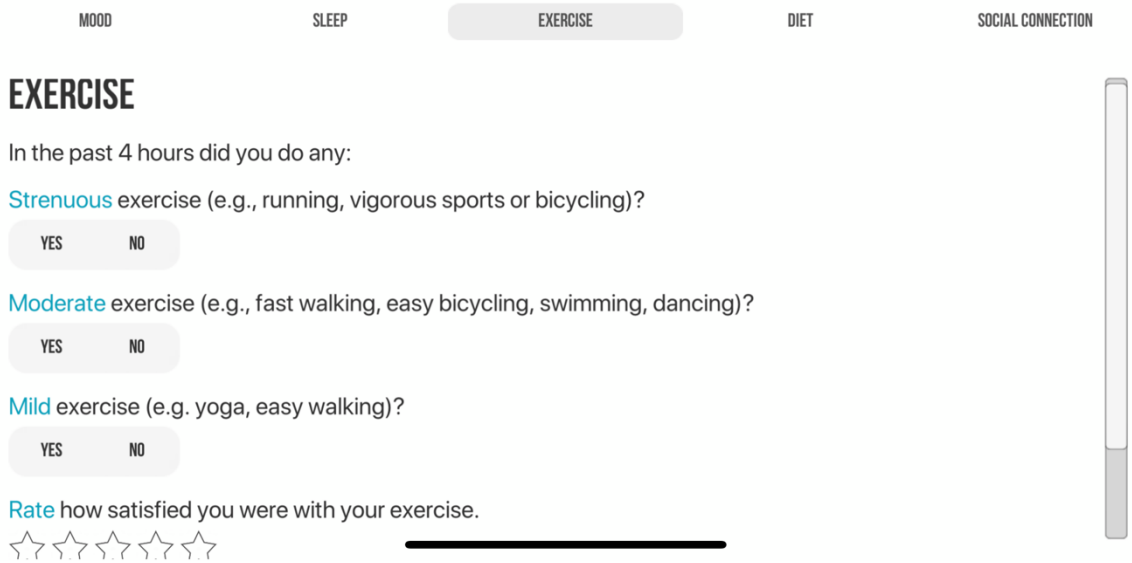

Figure 2. Self-report screen for exercise as participants see it.

preparatory sessions in which the laboratory's principal investigators (PIs) prepare guides to meet with participants. The present chapter relies on a distinction between these two interactional contexts, one of which is features mainly open-ended discussion between guide and participant, the other of which features mainly instruction from PIs on how guides should interact with participants. In PWP guidance sessions, guides ask participants for their own accounts of how they decide on self-report data, and about the activities that these reports describe. In preparatory lab meetings, the PIs focus on the ML system outputs, rather than asking participants themselves. Guides, who are present in both contexts, are shown to play a mediating role, interacting with PIs in one setting and participants in another.

Data-driven systems for health inevitably need to cope with paradoxical, missing or contradictory data. In the below cases, self-report data and passively sensed data conflict, and ML outputs imply unintuitive relationships. In these cases we see how repairing contradictions in data-driven diagnoses is attempted by PIs, who center the ML outputs and data visualizations, by guides who engage in dialog with participants about the issue, and by participants themselves.

In preparation sessions, which PIs orient toward data visualizations that they share using a screen-sharing feature of the video conference, these visualizations take precedence over human explanations of the causes of data anomalies and contradictions, even when a PI offers this kind of human explanation for consideration. Visualizations are shared as the center of the preparatory meeting and proposed as the ideal starting point for PWP guidance itself.

In the PWP guidance sessions, however, guides ask participants about their self-reports in an effort to address contradictions with passively sensed data. In this setting, we see one participant avoid the language of the self-report and EMA format in which it is solicited. Another participant, on the other hand, attempts to justify self-reports, using biomedical language and

quantifications of factors that could be relevant to the discrepant data. These PWP guidance sessions demonstrate how participants account for the process of "calibrating" their own sense of mild and moderate exercise with that of the data system, and to what extent participants attempt to accommodate the rhetoric and logic of the ML system as they understand it (as was shown in Chapter 2, Excerpt 2 for the PWP guide).

Broadly, guides are shown to play a mediating role. Because they are present in both contexts, it is up to the guides whether to forward theories, explanations and visualizations from the preparatory laboratory meeting once guidance begins. Also in the below case, in which a preparatory lab meeting occurs after session 1, the guide decides which comments from dialog with participants to share with PIs. In this way, guides play a key role in the practical maintenance of the separation of these two spheres, one of which appears in observation as dominated by data and prosaic visualizations of it, the other of which consists of open-ended interpersonal dialog and reflection on experiences relevant to mental health.

2. Literature review

There is a substantial existing social science literature on topics relevant to this chapter including how embodied and phenomenological experience is integrated into and expanded through digital health programs, how this kind of embodied knowing is reconciled with digital representations, how digital systems and their users take on agency and push against one another, and the challenges of contextualizing this process such that structural and environmental factors, affects, and sensory experiences are adequately accounted for.

In their critical discussion of digital phenotyping for mental health, Birk & Samuel describe self-report data like that involved in the present study as preferable to digital phenotyping systems that aim to establish patient mood on the basis of passive sensing alone, suggesting that "a person's interpretations of their situation – including how they think and feel

about their social life – is the central thing to gauge" (Birk & Samuel, 2020, p. 8). These authors are still cautious, however, about the prospect of using self-reports for mood, as they warn that digital phenotyping programs are often conceptualized around the data that digital devices can capture, and force self-report data to fit into that paradigm (2020, p. 11). In a study of digital self-tracking for fertility and contraception, Ellen Algera reports on some users finding visualizations of application data incongruous with their embodied experience, leading to "processes of continuously and iteratively comparing different types of data (measurements, cycle interpretations and experiences) and data sources (devices, self-checks and body awareness) and attempting to reconcile them" (2023, p. 255). In Algera's study, users often look to their embodied experience as the more trustworthy source and on this basis, develop criticisms of the digital system.

While digital health users often question and criticize data-driven insights on the basis of their embodied experience, some users learn new ways of sensing embodied experience as a result of using digital health technologies. In Lupton and Maslen's work on the subject the authors notice that, while some users were frustrated by the digital system when it failed to align with their experience, others used the system to learn to sense in news ways, "distinguishing 'false' bodily sensations from 'real,' discerning patterns, and enhanced sensory capabilities" (2018, p. 199). New ways of sensing were afforded by digital health and self tracking systems in other studies too, often by enriching the visual sense, allowing users to learn "not only to measure [a phenomenon] or predict it, but also … to see it more clearly and to examine it" (Kristensen & Ruckenstein, 2018, p. 3637); to "see the fertility in [the fluid]" (Algera, 2023, p. 250); or to "see meals in calories" (Lupton & Maslen, 2018, p. 197). Because of the way digital health tracking applications draw attention to the body, they not only allow users to "see" new

aspects of experience, but also develop their haptic sense of what Lupton calls the "felt body," disrupting the supremacy of the optical faculties (Lupton, 2013a, p. 398).

These inquiries do not simply compare the outputs of digital health systems and embodied experiences themselves, but also show how the two influence one another, sometimes forming feedback loops (Algera, 2023, p. 254; Ruckenstein, 2014b, p. 70). Lupton and Maslen use the term "data sense" to refer to the faculty that emerges as data-driven systems inspire new ways of sensing the body and embodied sensations bring about new uses of data (Lupton & Maslen, 2018; Maslen, 2017).

While the self-tracking program of the participants in Kristensen and Ruckenstein's 2018 study is described as "non-coercive and voluntary," other research has drawn attention to the ways that data-driven systems exert discipline on users, and the ways users push back (Pols et al., 2019). In a 2014 study, for example, Minna Ruckenstein found that participants "expected data flows to act on them," hoping that the digital health intervention could "take control" and bring about positive changes (Ruckenstein, 2014b, p. 74). Izabella, a participant in Della Bianca's study of menstrual self-tracking, started the study with a willingness to delegate decisions about her body to the application, but became more resistant over time (2021, p. 9). In one study of digital telehealth application for pregnant patients, on the other hand, the digital data system gave patients a further sense of agency, inspiring them to challenge doctors and other medical staff to repair gaps in medical records (Winthereik & Langstrup, 2010).

Several of these studies have in common a reluctance to assume a neoliberal subject who uncritically conforms to the normative digital health program under the auspices of individual responsibility (Della Bianca, 2021; Kristensen & Ruckenstein, 2018; Pols et al., 2019). Olivia Banner's crip theoretical critique of digital phenotyping highlights the ways users and afro-

surrealist film makers repurpose psychiatric technologies to resist the dynamics of racial capitalism and "imagine alternate visions of rationality, affect, and care" (Banner, 2019, p. 4). The present study contributes to this literature, showing how participants use guidance sessions to push against and augment the data driven health intervention.

In each of these studies, the subject of phenomenological experience and the digital health system do not interact in a vacuum. The challenge of interpreting, analyzing and acting on flows of digital data occurs in real-world, material contexts. Minna Ruckenstein points out how data visualizations inevitably abstract away contextual factors. Participants in this study become frustrated when the data system seems oblivious to the complex social relations that form the context of their health data (Ruckenstein, 2014b, pp. 78–79). Data visualizations and recommendations inevitably fail to capture contingencies of material infrastructures and complex social systems, causing gaps and breakages to emerge as digital health systems are implemented (Pink et al., 2018). This literature is relevant to the present study which focuses on how lab members and participants cope with contingencies like environmental factors and interactions between health areas in their production of self-reports.

3. Approach

The present chapter is based on observations of laboratory meetings and PWP guidance sessions. As I observed lab interactions, I also took field notes, using thematic tags to indicate which interactions relate to foci including participant self-reports of mood and exercise and lab members' efforts to understand how participants make these reports to align them with sensor data. I used these field notes to select relevant interactions for analysis. Other studies on these topics tend to implement interviewing and sometimes participant journaling to elicit phenomenological reflections on embodied experience (see eg. Algera, 2023; Ruckenstein, 2014b). In our case, the more naturalistic setting of laboratory meetings and guidance sessions

offers access to these kinds of participant reflections, which arise as part of the trial itself. This constitutes a unique aspect of the research design and serves as a starting point for further integrating participant reflection into digital psychiatry interventions (Grosjean et al., 2021).

This chapter is based on observations of interactions that occur in two distinct situations. First is the guidance preparation session. The planned purpose of this session is to prepare the guide for a particular participant. This includes reviewing visualizations and statistical summaries of the sensor and self-report data collected for the participant, and establishing or confirming which of the four personalized wellness plans (or PWPs), diet, sleep, exercise or positivity, will be recommended. Guidance preparation meetings include the guide who will be meeting with the participant, at least one of the lab PIs, and sometimes other guides. While guidance preparation meetings usually occur shortly before the initial guidance session, in some cases, guidance preparation meetings can take place immediately after the first meeting.

The second observed scenario is the guidance session itself. The first and sometimes second meeting is the time for the guide to explain the chosen plan and the rationale behind its selection. Subsequent sessions are described as opportunities for the guide to confirm that the participant is effectively implementing the recommended plan, track participant progress and address questions and concerns. Using field notes and thematic tags, interactions from two series of one-on-one guidance meetings involving two participants are selected from the observational data. For each participant, both the guidance preparation meeting and the PWP guidance sessions themselves are analyzed. Observational data are analyzed in combination with technical traces, including visualizations and spreadsheets that constitute machine learning system outputs (Geiger & Ribes, 2011). As will become clear, these technical traces are necessary for understanding the guidance preparations and guidance practices themselves.

Analytical procedures start with the selection of interactions from both guidance sessions and the laboratory meetings in which guides prepare for these initial sessions and finalize PWP selection. Transcripts from these interactions are then annotated for their relevance to issues around reconciling self-reports and sensor data. Sections of the transcript that include interactions which justify or challenge participant self-reporting, or bring these self-reports into alignment with sensor data are presented in their entirety below. By attending to this process in detail, the analysis sheds light on how participants communicate the complexities of their embodied experience and how, as they learn about the ML system and its outputs, they attune their self-reports and reflections in an effort to align with the data system.

Rather than serving as representations of the experiences of participants, or the attitudes of scientists, interactions constitute the process by which self-reports and sensor data become practically related in the course of implementing the intervention. Interactions themselves, not just the experiences and protocols referred to in these interactions, are shown to shape the outcome of the trial for both participants and laboratory members (Alač, 2020; Heath & Hindmarsh, 2002).

4. Observed Interactions

The two participants whose PWP guidance processes are presented here share several features. Both are assigned the exercise plan on the basis of machine learning outputs that indicate exercise-related features as those most directly correlated with changes in mood. Both participants are explicitly asked to help align their self-reports of mild and moderate exercise with passively sensed wearable data such as heart rate, step count and step speed. In both cases, discrepancies between these two measurements of activity become problematic for laboratory members as they discuss the PWP plan recommendation. Also in both cases, the participant brings up complicating factors that are not accounted for in the protocol in the trial. For both

participants, this involves interaction between exercise factors and anxiety and social interaction factors, the latter of which fall under the purview of the positivity plan, which is distinct from the exercise plan. Each in their own way, the two participants share the complexities of their attempts to implement the plan in the real world, revealing the gaps and breakages in the data driven system, which aims to collect data that represents the totality of user behavior and mood.

4.1 Preparatory meeting for Participant 1

The PWP guidance process for Participant 1 begins once an adequate number of data points have been accumulated, both "passively," through sensors on the participant's smartphone and wearable device, and through participant self-reports. Once these data have been collected, a PWP guide is assigned, and the PI arranges to meet with the guide to discuss ML outputs derived from participant data. This preparatory meeting serves to prepare the guide to begin PWP guidance sessions, to elucidate ML outputs, confirm which PWP plan recommendation these data best support, and strategize how to communicate these findings to the participant. The meeting includes the two laboratory PIs, two guides, one of whom had already been assigned to the participant under discussion, and me, the observer.

Starting in the guidance preparation meeting, implementing a PWP for Participant 1 requires combining self-report data, which the participant provided as a subjective account of activity and mood, with data on step count, step speed and heart rate that was automatically collected by sensors on the participant's wearable device and smartphone. As is shown below in Excerpt 1, making sense of the data that are based on self-reported durations of mild and moderate exercise elicits a human interpretation by one of the lab's PIs. Through this interpretation process, one of PIs, here called "PI2," develops a theory of participant behavior that explains the seemingly contradictory data. This interpretation is said to be too technical to be

worth sharing with the participant—it is a kind of interpretive knowledge that is specific to the domain of psychiatric scientists and is presumed to be of no interest to the participant.

Excerpt 1. PWP guidance preparatory meeting for Participant 1

PI1: ((Opens Spearman figure (Fig. 1)))So this is a plot of whether things positively or negatively relate, which you may or may not want to see. Um, so, so, the more positive here is, depression is more positive. So actually the, like we saw, more mild exercise is not actually that good for the person but more, um, when they're doing moderate exercise is actually good. So um so consider, uh.

PI2: And this is just univariate correlations. This is not a multivariate.

PI1: yeah these univariate mea::, yeah, one at a time these have been modeled, with, just to help you with the sign if you ever care. I don't think as, um, as a person who's discussing the plan with the individual you actually would go into these relationships. Um. I think the bar graphs do a good enough job. Um. I'm not going to go into more details but essentially showing the same thing as the bar graph. That, you know when they were doing more exercise, more milder exercise, they were positive on the X axis is more depressed. Um. Negative on the X axis is less depressed or more well-being. So when they were doing moderate exercise they were less depressed and had greater well-being. So that's why. It's a negative relationship with moderate exercise but positive relation mild exercise. So I believe that this is not really that helpful for your meeting. ((Closes Spearman figure (Fig. 1))) It's just there. Um. This table, ((opens Intervention table (Fig. 2))) uh, will give you, um, just you know just, uh, you know, the, the values that you saw plotted in the bar graphs. Um. They are, the actual values underlying them, are in here if you're ever interested in that. Um. So you only saw the summary values on the top and that's what's here. But we also do, like, how many counts were there in each bin when they were more or less, um, uh depressed. Remember the other thing is that the way we've shown this is that we're binning. The, the way it's shown is bins. Like, more depressed bin or less depressed bin. But the

model's actually continuous model. This is just for interpretation purposes. Um. There are bins in here ((closes Intervention table)). Um, you, so, I think your discussion with the participant is really around this main figure ((opens Shap figure (Fig. 2))) which is that, the SHAP figure. I think mostly this one. Um. Another figure that does, that may or may not be interesting to show,((closes Shap figure. Opens Shapely fi figure (Fig. 3))) is, is, is this one, is the Shapely fi figure, which is basically showing you, um, the importance of the features as they come out from the model. But they're, uh, here there's no sign or anything like that. In that, uh, for this person, it's just showing the ranks of what's most important. So mild exercise is most important, then breath response, um, anxious mood, uh, past day exercise, moderate exercise, then sleep satisfaction and so on. So just ranking the variables. Um.

PI2: This can be helpful when we're trying to identify what the optimal plan is. Because you might look at the top five but really the first one is the most highly predictive, or the second one, first two, or something.

PI1: Right.

PI2: This seems kind of more typical for like what you generally shown me. Which is, there's a drop off, but, you know the first five are seeing relative. This, this seems like a pretty typical drop off

PI1: Yeah so it'll be a typical, like you're saying, a typical drop off. But at the same time we find that, maybe, would you say from looking at it, that it's more important to reduce mild exercise than to enhance moderate or. I think they kind of go together. Umm.

PI2: I think that's what's the human interpretation which is what I would make as well. And where this is, like, very tricky to just give an automated AI-based thing, you know, without our interpretation. Like somebody would say, "oh stop exercising in a mild way." My interpretation is maybe this person is either doing mild or moderate exercise most days. And the days they do the moderate they're doing better. That would be my interpretation. The mild exercise is not, maybe, hurting but it's, like, the days that they're not doing the moderate is when they're doing the

mild, and that's why it has this positive relationship. That would be, yeah, that's.

PI1: ((closes Shapely fi figure (Fig. 3))) Um. Great. So you guys don't have to go into the extra folder. These things are, you know, when we publish this, these are extra figures that are useful. But the by domain folder may be useful to you. What we've done over here is that for each domain that we're mapping, so in this case the exercise domain is of interest to you, um, you can just look at all the features that come in the exercise family, basically. So, um, ((opens Exercise spreadsheet(Fig. 4))) if I were to open exercise, and there's no, um, figure for this, but in general what you'll see here is, um the rankings are still the same. Past day mild exercise had the top ranking; past the moderate exercise had the next, um, ranking in this domain. Um, but essentially what we're seeing here is, um, you know, they're, they're exercising a little bit, uh, when they, they have reduced mild exercise and greater, um, moderate exercise they're doing better. Um, that we know. Um. There're other actionable things, um, that might be there that you could tell from here, or. Like what [PI2] was mentioning a little bit about calories. So, um, it's, like, natural that when you're doing more moderate exercise you would burn a little bit more step calories. So you're burning, um, you know, on average 20 or so more calories on this day. So if that's something that they still have a watch, and it's something that they look at on the watch, and step calories is a feature that they can look at, this is something might be of interest to them. Um. This is, uh, you know, meters so the cumulative step distance that they walked in the past um, 12 hours, um, relative to whenever they were giving a rating. Um, so they've walked more in this case. So, uh, so this is, this is cumulative over, so step distance is always cumulative over the last 12 hours. So 2.89 kilometers is like, you know, uh, or just like, my, I think that's like, a little bit, a little bit like about two miles or so, is about how much they're walking in the, in the, in the, in about a 12-hour time window. And that's, that seems natural. It's not like at that point when they're feeling, um, better mood is when they've, not just at that point, it's uh, like. What you need to understand is that it's in the past 12

hours. So walking about two miles in the past 12 hours makes sense. So here they walked a little bit less. Um. Again, exercise calories are a little bit more. Um. So these are watch driven features. Exercise duration is a little bit more, in terms of what's coming out of the watch. Um. Step count is more, um. Heart rate's about the same. Um, satisfaction from exercise's just slightly more. Uh, not, not you know, just like a point more uh, than, than previous.

Overall, the above excerpt is of interest because it demonstrates the process by which self-report data and automatically collected sensor data are made compatible to contribute to a behavioral model of mood. The two PIs of the laboratory contribute to this reconciliation process in different ways. PI2 suggests a "human interpretation" of self-report data to explain how it relates to the rest of the dataset. PI1 draws attention to the sensor data that align with self-reports, emphasizing the "natural" relationships that exist between these two kinds of data.

In the guidance preparation discussion excerpted as Excerpt 1, PI1 starts by presenting the Spearman correlation plot for Participant 1 (Fig. 1), which shows how self-report and sensor variables correlate with mood. A unique feature of this plot is its inclusion of what PI1 calls "the sign"—depending on the direction that each bar extends from the centerline of the plot, it may represent either a positive or negative relationship with depression score. Positive sign, in this

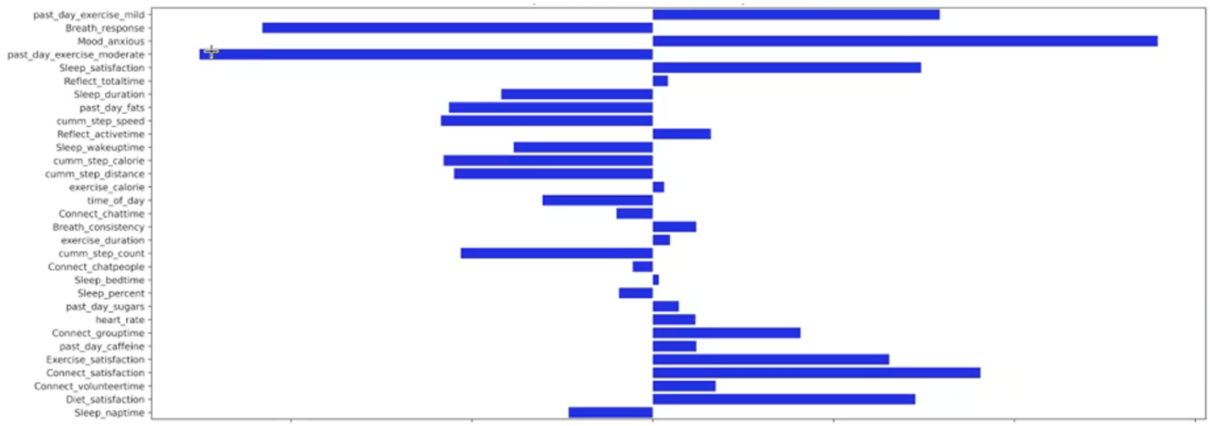

Figure 3. Spearman figure. Positive or negative correlation is shown on the X-axis. Factors are listed on the Y-axis.

case, indicates a correlation with more depression, which PI1 describes as synonymous with greater well-being. Likely because of their unexpected and contradictory nature, PI1 explicates the Spearman values of mild and moderate exercise duration, highlighting that longer durations of mild exercise seem to be correlated with worse mood, and longer durations of moderate exercise with better mood. The Spearman plot displays this difference in a pronounced manner compared to the figures discussed below.

The first response from PI2 is to confirm that the plot represents univariate correlations. The PI wants to be sure that each variable is shown according to how much variation in mood is explained by variation in that variable alone, without any adjustment for other variables that may interact with this relationship. This question sets up PI2's later explanation for the seemingly contradictory indication that mild exercise correlates with higher depression scores.

PI1 continues the discussion by explaining that the Spearman correlation plot is not appropriate for showing participants during PWP guidance sessions. PI1 then explains again that mild and moderate exercise seem to have opposite relationships with mood, and then reiterates

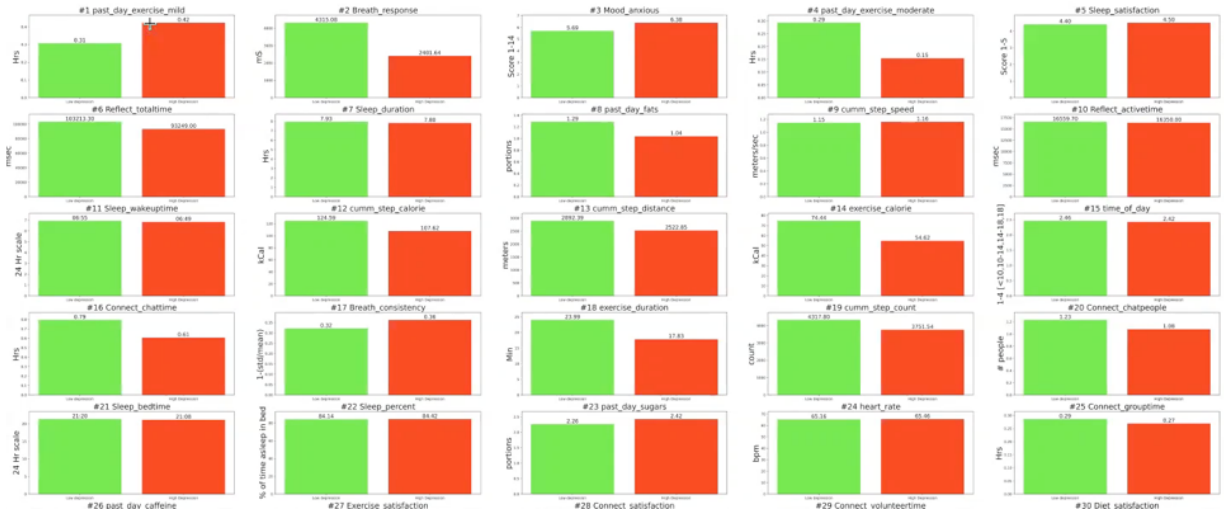

Figure 4. Shap figure. Each factor is shown as two "bins," each marked with the depression score. The Y-Axis represents quantity of that factor

that this plot is not “really that helpful for your meeting.” The PI goes on to suggest that the Shap figure (Fig. 2) should serve as the “main figure” in conversations with participants. The Shap figure shows each recorded variable in a box that contains a bar plot with two bars. The green bar represented positive mood and the red bar represents negative mood. The bars are sized and labeled to show the quantity of the variable that correlates with either positive or negative mood. The PI only briefly displays the Shap on screen before switching to discussion of another plot, Shapely Fi, which is also said to be appropriate for guidance meetings. The Shapely Fi plot shows a list of the variables that are highly correlated with mood which are, like in the Shap figure, ranked by the importance as a feature of the model. PI1 highlights that there is no sign on this plot, making it impossible to see that mild and moderate exercise are contributing to higher

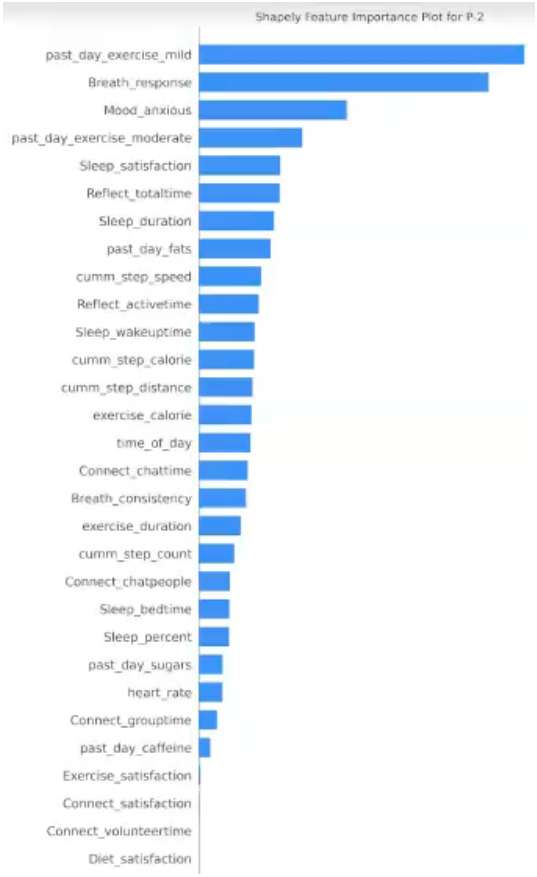

Figure 5. Shapely fi figure. Factors are ranked by correlation which is shown on the X-axis.

and lower depression scores, respectively, instead simply showing that both are relevant variables.

13	cumm_step_distance	MEAN	2522.85	2892.386	
	meters	COUNT	26	35	
		STD	2562.127	2782.418	
		CI95	1467.491 - 3578.210	1922.638 - 3862.133	
14	exercise_calorie	MEAN	54.618	74.445	
	kCal	COUNT	26	35	
		STD	63.123	114.598	
		CI95	28.617 - 80.618	34.505 - 114.386	
18	exercise_duration	MEAN	17.829	23.988	
	Min	COUNT	26	35	
		STD	20.588	36.034	
		CI95	9.349 - 26.310	11.429 - 36.547	
19	cumm_step_count	MEAN	3751.538	4317.8	
	count	COUNT	26	35	
		STD	3748.978	4111.555	
		CI95	2207.306 - 5295.771	2884.812 - 5750.788	
24	heart_rate	MEAN	65.456	65.161	
	bpm	COUNT	26	35	
		STD	2.677	3.228	
		CI95	64.353 - 66.558	64.036 - 66.286	
27	Exercise_satisfaction	MEAN	4.769	329	
	Score 1-5	COUNT	26	35	
		STD	0.639	0.56	
		CI95	4.506 - 5.032	4.633 - 5.024	
	past_day_exercise_strenuous	MEAN	0	0	

Figure 6. Exercise spreadsheet showing the statistics underlying the above figures.

PI1 then asks the other PI for thoughts on the relationship between and comparative importance of the mild and moderate exercise variables. PI2 suggests, in what he calls “a human explanation,” that the participant is likely reporting either mild or moderate exercise each day. The statistics that indicate that mild exercise correlates with lower mood should not be interpreted as a simple causal relationship in which mild exercise causes depressed mood, but rather as a trade-off effect—the lack of moderate exercise is what causes the depressed mood, which corresponds to days that also are reported to include mild exercise. This PI highlights the fact that this explanation could not be produced as part of any of the ML outputs, lamenting that in this case it is “very tricky to just give an automated AI-based thing, you know, without our interpretation” (lines 64-65).

PI1 does not address this theory directly, but instead responds by closing the Shapely FI figure and shifting the conversation to another document, the Exercise Spreadsheet (Fig. 4), which the PI displays through ZOOM screen sharing functionality. This transition into discussion of a different document constitutes an interactional affordance of the remote, video conferencing format, which allows the participant who is sharing their screen to guide the visual context of the interaction.

PI1 takes advantage of the multimodal affordances of the teleconferencing platform by highlighting individual cells in the spreadsheet that is displayed as a shared screen in the video conference. The PI selects cells in the sheet, highlighting variables for which the "higher mood" bin contains activities with more substantial exercise content. In this way, the PI supports the notion that moderate exercise is preferable to mild exercise: "it's, like, natural that when you're doing more moderate exercise you would burn a little bit more step calories."

The interaction excerpted in Excerpt 1 includes several insights regarding the process of reconciling self-report data and sensor data. In this case, the process of reconciling these data is challenging because of the unexpected positive relationship between mild exercise duration and depressed mood score. Each PI take their own approach to tackling this issue. PI2 suggests that this is a case in which human interpretation is required to contextualize the statistics and avoid erroneously attributing depression to greater mild exercise duration. PI2 develops a theory about how the mild and moderate variables could be interacting so as to produce what looks like a causal relationship between mild exercise and poor mood.

PI1 takes a different approach to describing the data, drawing attention to plots that place less emphasis on the sign, and highlighting the sensor data that seem to accord with a general

benefit of more strenuous exercise. PI1 also repeatedly suggests that the Spearman correlation plot, which most clearly displays the divergence of mild and moderate exercise factors, is not appropriate for PWP guidance sessions with the participant, suggesting that in interactions with the participant, knowledge of the sign is unessential, supplementary information.

The controversial argument that big data and ML driven technologies have led to a "end of theory" is difficult to defend partly because of the diverse range of different projects that use big data and AI, each of which involves different kinds of interaction and communication. In this case, we see the particular interactional strategies involved in a conflict between a theoretical explanation and a more prosaic appeal to additional data of different types. PI1 selectively displays and emphasizes data visualizations that impart a general idea of more strenuous exercise as being beneficial, while avoiding the Spearman correlation plot which emphasizes the difference in sign. PI2, on the other hand, develops a theory to explain, in terms of the deeper cause or what is called in psychiatry, the etiology, of the surprising data.

Preparation of the co-present guides, in this case, includes exposure to these two different models of how they could explain ML outputs to participants. The strength of PI1's appeal to data as they are visualized, rather than a conceptual etiology, is enabled by the visualizations themselves, which come in various versions, some of which show the "sign" and some of which do not, and the ability to choose and share these visualizations by sharing the screen. The separation of the sign from the quantity is a theoretical position that is not explained verbally in this interaction, but rather built into the structure of the plots that PI1 shares. The guides who are being trained in this interaction mediate between the data scientists, PIs and participants. In this case, the role of the guide is modeled by PI1 as following from the plots, which contain their own conceptual structures and can therefore be selectively shared to emphasize some aspect or

interpretation of data over others, without the need for a verbal argument or explanation. This largely accords with guidance trainings as they are described in chapter 4—the role of guides is not to investigate or raise questions about the design or implementation of the trial, but rather to simply say the right things to the participant to help activate their intrinsic motivation.

As Minna Ruckenstein put it in her self-tracking study that also involved "coaches" similar to the PWP guides, "data visualizations operate as shared artifacts for co-constructing and negotiating meaning" (Ruckenstein, 2014b, pp. 76–77). In other contexts, too, scientific data visualizations can constitute a shared space for multimodal semiotic processes whereby interlocutors work together, using gesture and indexical language, to make meaning and carry out laboratory tasks (Alac, 2011; Vertesi, 2012). In this case, however, the shared screen is not used as a shared semiotic space, but rather as a container of self-evident data-based insights. The strength of PI1's account of the situation comes from their role as screen-sharer, a position which allows them to determine and interchange the object of joint attention, be it a chart or a spreadsheet, which forms the locus of indexicality, "what we're seeing here." By rapidly switching out this object of joint attention over the course of a single turn, the other participants have little chance to engage with the figures using gesture, indexical language or otherwise. In this way, the video conference scenario with a single screen-sharer does not invite collaborative meaning-making, but rather privileges a single sharer and the content they have prepared. Similarly to the PWP trainings described in chapter 4, this preparatory session consists of the PI showing the guide what to show the participant, rather than collaboratively coming to an understanding of the meaning of the visuals.

4.2 PWP guidance for Participant 1 begins

When guidance sessions begin, they feature a continued interest in the challenge of reconciling self-report and sensor data. After having assigned and explained the plan to the participant in

session 1, the guide meets with Participant 1 to ask about the exercise plan that was suggested in their first meeting. In this and the following meeting, getting a sense for how the participant selects mild and moderate exercise categories is a topic of conversation. While PI2 explains the contradictory exercise findings according to a trade-off relationship not captured by the algorithm, and PI1 explains it by highlighting visualizations of sensor data showing the benefit of strenuous exercise, the guide demonstrates yet another way of investigating the contradiction, which is based on eliciting experiential reflections from the participant.

In session 2, the guide starts by asking the participant about how they changed their behaviors since beginning the intervention, and how they feel after they complete the exercises that are recommended as part of the PWP. The guide prompts the participant to reflect on these exercise experiences. After having solicited these reflections, the guide explains that they could help explain the mild and moderate exercise data.

Excerpt 2. PWP guidance session 2 for Participant 1

P1: I did exercise. I thought about what makes me happy when I'm physical. I decided that I prefer to be in the garden, in the yard. So rather than at a class I decided that I would just spend time outside. And play in the dirt.

G1: Well that's great that you found the time to reflect over the last week.

P1: If the data shows that I'm happiest when I'm physical I figured that would be the best place to be.

G1: I didn't see any gardening recorded in the log you sent me. Is that on top of what you sent me?

P1: So when I started processing what we talked about last week, I did my usual kind of exercise, I might do it and I might be happy afterwards but I'm not sure I'm happy while I'm doing it. When I'm in classes I watch the time and. I watch the steps. What the pedometer's doing. When I'm outside I don't watch the time I'm just. There. I just spend time out there and it's just mindless nice a happy place for me.

G1: That’s amazing that you were able to notice that about yourself and have such a positive response to gardening. When you’re on these walks it sounds like you’re not enjoying it as much as the gardening. How are you feeling after.

P1: I feel good. But it’s not like I’m enjoying them while I’m doing them.

G1: There is a reward in doing something you, you didn’t initially want to do. Can be very stimulating for the brain

P1: I know health-wise it’s good for me to do it. But the gardening is just nice and fun and doesn’t have any of the other stuff around it.

G1: It seems like right now there are two options. I would hesitate to substitute everything you’re doing for the gardening.

P1: No I wouldn’t do that because the other stuff is much more physical than the gardening. I plan to continue to do the walking and all of that and take the classes. But I just want to make sure that I’m making the time to get out and work in the yard.

G1: That’s amazing. That’s what I was gonna say. You seem to have a really good sense of self. Let’s continue the gardening. Does that sound like a good plan for you right now?

P1: Yes

… …

G1: Tell me a little bit more about the fitness classes.

P1: Well this one yesterday was this restorative yoga and it think, I was watching the clock because I was meeting my son and grandson afterwards. I really appreciate the weight part because I think you know the walk is weight bearing. But I’m not lifting weights so, you know, I think lifting the weights as part of the class is good for my body.

G1: Absolutely. And you know there’s kind of the two. Aerobic exercise, walking, running, cycling. And then the resistant training. And there’s a lot of research that shows that we need both of those.

P1: She also does balance and I think that’s really important.

G1: How do you feel, kind of, the next morning. Do you feel any difference in your mood or energy?

P1: You know I haven’t really paid attention to that. I live in a multi-story house. I do notice that the next day I’m sore. Not sore enough

53 that I'm taking Advil or anything but. I notice it. You know I
54 haven't been paying attention to my mood the next day.
55 G1: Sometimes it can be hard to tell just sort of day-to-day. Because
56 the initial data was suggesting a difference between the mild and
57 the moderate exercise. Just kind of trying to compare how you're
58 feeling the next morning. And taking in particular account of your
59 mood and how you're feeling.

In this excerpt, the guide investigates the various forms of exercise in which the participant has taken part since their last meeting. The participant starts by sharing that gardening is what brings happiness, and so, in combination with the PWP plan recommendation which seems to suggest that "I'm happiest when I'm physical," the participant feels justified in spending this time "in the dirt " (lines 1-3). When the guide asks about how the gardening was logged into the application as a self-report, the participant responds by shifting the conversation back to a holistic account of the experience of different kinds of exercise, explaining how gardening brings joy and immersion in the moment—experiential aspects that are absent from the exercise class experience (lines 9-14). Later in the conversation, the guide asks for more details about the exercise classes. The participant responds by reiterating how she experiences an awareness of the slow passing of time (lines 35-36). The guide asks about how the participant feels after exercise, and the participant shares that soreness the next day has occurred.

Throughout the excerpted interaction, the guide attempts to gauge how the participant feels after exercise. The guide eventually explains that these questions come from his inquiry into data that is "suggesting a difference between the mild and the moderate exercise" (line 51-52). Given the context of the preparatory meeting, it is likely that the divergence in how mild and moderate exercise are shown to relate to mood in the ML analysis informs these lines of questioning, which the guide continues in the third guidance session to ensure that the "data is, kind of, lining up with the same thing that [the participant is] experiencing." In both sessions, the

guide pursues an informal, post-hoc kind of momentary assessment of mood in the hours following exercise by asking, “how are you feeling after?” and, “how do you feel, kind of, the next morning?” This approach accords with the structure of the self-report system, which is based on ecological momentary assessments or “EMAs” which represent mood experienced in a single moment. Guides attempt to mirror the logic of the technology that they aim to supplement in a phenomenon also described in Chapter 2, as well as other studies in digital psychiatry (Semel, 2021).

The participant’s accounts of exercise experiences, on the other hand, conflict with this momentary understanding of mood—to the participant, the passage of time feels different in different experiential contexts (lines 11-12). The participant emphasizes how exercise experience extend over time, rather than the feelings they elicit in particular moments. They also distinguish between feeling good in the process of doing something and feeling good after. By the end of the excerpt, the guide has acknowledged that the participant has reflected on exercise and mood. However, the guide also accepts that these reflections will not yield an EMA kind of data that could resolve the question of mild exercise correlating to low mood, telling the participant on line 55 “it can be hard to tell.” The participant’s experiential account of how different activities offer different kinds of fulfillment and seem to occur over time in different ways is incompatible with the logic of the EMA. PWP guidance work, then, includes the challenge of to bringing into alignment these distinct ways of conceptualizing exercise and mood. Pointing out the problems with attempts to live a “data-driven” lifestyle using self-tracking technologies, Natasha Schull uses a similar experience to the “play in the dirt” described by Participant 1 as her example: “One imagines a future iteration of the Fitbit advertisement in which the protagonist who gives herself over to carefree backyard play will not only trust her suite of devices and software to

remotely track her but also to keep her on track, interrupting the flow of her experience to prompt her – when an algorithmic analysis of her own real-time data deems it necessary – to eat, drink, or rest" (Schüll, 2016, p. 14). Here, Schüll draws attention to the irony of attempting to integrate a data-driven understanding of mood which, in this case is limited to the logic of the EMA, to the flow of experiences in which we engage the people and things in our environments.

The PIs and the guide each take their own approach to the unintuitive correlation between mild exercise and lower mood. PI2 proposes a theoretical explanation of the unexpected ML output, and PI1 instead refocuses the conversation on other more robust seeming variables and the visualizations of these variables. The guide, on the other hand, addresses this issue in the guidance meeting itself by asking the participant how she feels after exercise, soliciting a kind of post-hoc EMA. The communicative disjunction shown in Excerpt 2 demonstrates an instance in which simply asking for further subjective accounts of experience cannot disambiguate or resolve the breakages and paradoxes of ML findings which are designed to make subjective accounts and sensor data intelligible side-by-side. The experiential reflection in Excerpt 2 is the result of the dialogic, interpretive nature of the sessions, which contradicts the conceptualization of the trial as being mostly free of human interpretations, and of PWP guidance as fundamentally distinct from talk therapy because of its basis in data and the empowered individual participant (See Chapters 2 and 3 for more on this conceptualization).

While this disjunction might lead us to presume that the interactions of PWP guidance end up having little influence on the course of the intervention, as they are not deemed valuable in formal accounts of how the trial works, the case of Participant 2 demonstrates how PWP guidance interactions can cause a change of course.

4.3 Participant 2 begins PWP guidance in advance of preparatory meeting

For Participant 2, the initiation of PWP guidance proceeded in a different sequence of events. The guide conducted the initial meeting before having had an opportunity to meet with the PIs for a full-fledged preparatory meeting, relying instead on brief exchanges confirming that the exercise plan should be selected. Immediately after this first meeting, the guide met with PIs for a preparatory meeting to discuss how the PWP guidance process should proceed.

The first PWP guidance session, which occurred in advance of a preparatory meeting of PIs and guides, includes extensive discussion of "exercise plan" and "positivity plan" related factors including social connection and anxiety. An excerpt from the meeting in which these topics are discussed is presented as Excerpt 3.

Excerpt 3. PWP guidance session 1 for Participant 2

G2: So as you can see the most two things that are affecting your mood is, one. Workout. The distance you go the intensity of the workout that you do. And the second thing is connection and gratitude, which means a sports that you do and setting of groups is basically helping you more. So before we go more into that why don't you tell me a little bit about your workout. How much do you workout? How do you workout?

P2: Um usually it's alone just cause I don't really like having distractions. I also get really self-conscious when there's other people around. I don't, there's just, I think a lot of people experience that when they're first really getting back into working out as you feel like everyone's looking at you even if they aren't. Like I know that it's not really probably what's going on, but it's, you know, insecurity can affect how good your workouts are. So I just choose not to put myself in that position and kind of do it myself. So I like to start by doing like a walk in the neighborhood with my dog and then I have a rowing machine and a spin bike and a set of weights in my garage which is usually where I do my workouts. I tend to do, like, like low volume weight bearing exercises just 'cause doing like really high intensity workouts makes me kind of anxious. I think the increase, the rapid increase in your heart

rate, isn't the best workout for some people I'm one of those people. So I prefer to do kind of like slow movements, really. I used to work as a PT intern so just trying to figure out ways that I can incorporate as many of my muscle groups as I can and, kind of, doing that. But, yeah, usually alone.

G2: So what would you define as mild workout and what would you define as moderate workout?

P2: Um. Mild I would say probably floor work. So I'll get out like the yoga mat and I'll do you know like crunches or planks. Kind of, things that aren't really involving too many groups at once. Moderate would probably be my rowing machine 'cause it's using both my legs and pulling back with my arms and I'm going at a much faster rate than I'm doing for work.

G2: OK and you identified these as mild and moderate in your [mobile application] when we were gathering data, is that right?

P2: Yeah. I kind of had, I categorize things in my brain before I decide how I'm going to answer them,

G2: It's absolutely OK. I just need to know what categories that you put on my end, what you put on moderate, because we're gonna try to move from mild to moderate a little bit more. That's what they get is showing. So I wanna know, like, how your workout routine is. How many minutes you put a mile what would you define as mild? Gow many minutes to moderate? And then from there we can slowly start to raise the amount of moderate for you from mild and see by the end of the six weeks if we can get to that 150 minutes a week. do you do

P2: 150 minutes a week?

G2: 150?

P2: yes so like 20 something minutes today I know on even I can't match

G2: that

P2: Yeah. I was like that's, that's really low. Yes.

G2: OK OK. So, uh how much of it is mild how much of it is moderate.

P2: Let's see. I would say probably almost a 50/50 split probably. How many minutes would you say is the mild how many the moderate I would say on average a day I like to get an hour in whether it's you know going the extra step to walk 15-20 minutes to my car, when

P2: I'm leaving work. I would consider that probably the, the mild. And then for the moderate I would try to meet that 30 minutes minimum.

G2: And you do rowing for 30 minutes minimum?

```
59      Yeah. Got a built in fan you can you can just keep going really.
60      It's pretty, it's great.
61  G2: And you don't raise your heart rate at all?
62  P2: With the rowing it does. Um. But I tend to focus on, like, slower
63      precise movements. A lot of people like to kind of ramp it up, by.
64      So what I would do is increase the resistance so that when I'm doing
65      each movement there's a little bit more I'm working against as
66      opposed to doing low resistance more reps. So I would do higher
67      resistance less reps.
```

The excerpt begins after the PWP guide has shown the Shap figure to the participant and summarized it in terms of workout (elsewhere called "exercise") and social connection (elsewhere called "positivity" and "gratitude"), the PWP plans of highly mood-correlated factors. As was the case in Excerpt 2, the guide then asks the participant to elaborate on experiences that relate to these kinds of behavior. In response to the guide's inquiry regarding exercise, the participant explains that the presence of other people during exercise is distracting. The participant experiences self-conscious feelings when working out around others. The participant uses a second person voicing in phrases like, "everyone's looking at you" (lines 7-13), which, in combination with the comment that "a lot of people feel that [way]" serves to frame the account in terms of general experiences that are common to general populations. Still making use of the second person, the participant goes onto explain that they limit the intensity of their workouts to avoid anxious sensations that, for them, go along with "the rapid increase in your heartrate." The participant also brings up the number of muscle groups involved in the activities, reporting an effort to use as many muscle groups as possible while avoiding anxious feelings. The participant reports an increased awareness of which muscle groups are involved in exercise activities as a result of experience working as a physical therapy (PT) intern.

The quantity of involved muscle groups comes up again as the guide asks about how the participant distinguishes between mild and moderate exercise. The participant explains that

moderate exercises use more muscle groups as well as "a much faster rate." After the guide confirms that the participant is explaining the categories they choose when they complete self-reports in the mobile application, the participant assures the guide that they "categorizing things in [their brain]" before deciding how to complete the self-report.

By this point in the trial, both the guide and the participant are concerned with the challenges of self-reporting exercise, as is made clear by the guides repeated inquiries about how the participant decides to select the "mild" or "moderate" exercise category, and the participant's explanation that they consciously conceptualize these categories in advance to ensure effective self-reporting. References to concepts from physical therapy such as heart rate, muscle groups, sets and reps, serve to legitimize this procedure further. Unlike Participant 1, this participant acknowledges the feasibility of categorizing exercise experiences and attempts to conduct and explain this categorization process in a manner that accords with the rhetoric of biomedical science.

The participant's theory of exercise, however, is complicated by the interaction of exercise factors, social factors, and anxiety, which is typically seen as a comorbidity that is common in depressed patients, but not a behavioral factor that can be grouped into one of the four families and used as the basis for recommending a plan. While physiological interactions between anxiety and exercise are not reflected in the structure of the intervention, the participant nonetheless frames these interactions in physiological terms, likely in an effort to be understood by clinicians including the guide. As the guide brings insights from this meeting into a preparatory meeting with the laboratory PIs, the interplay between anxiety and exercise factors, as well as the legitimacy or credibility of self-reports comes under new scrutiny.

4.4 Guidance preparation session after initial meeting with Participant 2

As was the case with Participant 1, the preparatory meeting regarding Participant 2 includes discussion of the limitations of the ML outputs for producing a viable PWP recommendation. For Participant 2, exercise factors including duration of mild exercise appear in the top factors determined by the ML algorithm. However, the algorithm output also shows that these relationships are not strong relative to other participants, yielding a high error statistic. Further, the mild exercise factor is once again paradoxically correlated with higher depression scores.

In an excerpt presented below as Excerpt 4, PI1 explains the high margin of error (as quantified in the statistic "Mean Absolute Percent Error" or "MAPE") of the ML outputs produced from Participant 2's data. As the PI responds to a data scientist's comment about the high MAPE for Participant 2, the guide interjects to share what he gleaned from the initial guidance meeting.

Excerpt 4. PWP guidance preparatory meeting for Participant 2

PI1: Yeah I mean, the person did not have great, the LSTM model was one of the worst models. They had like a 75% umm MAPE. So 50% was like for some like adaboost model, the A/B model, and that was OK. But anyway yeah I'm not sure what we might want to do specifically. But I just go into this person has had slightly greater error than other people. Don't know why.

G2: So for [Participant 2], I, I talked to her and she is just seeing anyone. Connectivity is very low. I encourage her to join a gym or go to a group workout. The other thing she mentioned was that her sleep is really bad. She goes to bed at 9:30 wakes up at 3:30 almost every night. So these are the stuff that I got from the interview from patients. But I don't know. If you want to do it, obviously, data-driven, then we have to have a data to support either one of these two plans. You can do a more detailed analysis of its sleep data that we have. For connectivity we can do that and change our plan. If not, I can keep forcing the 30 to 40 minutes workout and she said she's going to join a gym next week

and try to do a moderate workout with something. What do you think?

PI1: Sort of already bought into a moderate exercise plan, that's fine. We can go with that since that's one of the top factors. Um.

G2: Why do you think I should bring the exercise? Like it's 30 minutes. Do I want her to do an hour daily by the end of 6 weeks?

PI1: No, it's not really increasing, potentially, I mean, I think it's more like maintaining consistency, more, rather than increasing time so much. I mean, I think it's just that they're doing a lot of, I think, just, the data suggests a lot of mild exercise activity for long durations that is not helpful at all because that leaves room for rumination.

G2: The mild exercises are like walking to the car from the gym. So it's not that I can lower their mild exercise. They just put their daily living as their mild exercise.

PI1: Well it’s like an hour of walking the day. Like the cumulative step this day for them is like, not an hour, uh a kilometer, a kilometer or so meters. I mean if it's being read right then cumulative step distance of the kilometer per day versus.

G2: She said her daily living.

PI1: OK well is she. Um. Well what is it do you feel the exercise plan was for her or the one should switch to the positivity plan?

G2: She already does a lot of exercise but she has zero contact with other people. She's self-conscious so positivity is gonna be very good for her. That's, that's, that's very clear. Uh. But you know as if you want to do exercise, I could do exercise. But I can’t.

PI1: I mean, You already had a first session. But I can't, I was just trying to be curious about what you got buying for and if umm, you know, the participant ready to do a start with the, ehh positivity point, great. that sounds great

G2: OK perfect I'll go with positivity next session OK great thank.

P11: So usually [Guide 2], do you in the first section or is this something that you review after the first session, or not at the first session, or are you reviewing it after the first session?

G2: No, no. I decide them on the first session and I convinced them to buy in into it. [Participant 2] was a was a very different case. Like I didn't expect her to say, I already do 150 a week. I already do 30 minutes of rowing a day, Then I talked, well do you

wanna move your mild exercise to a little more moderate exercise, and she said "the mild exercise that I put in were just walking to my car from work and coming back there's nothing I can do about the mileage" so then I decided OK do you have a gym can we move your mild exercise to a more group setting mild exercise? And she said, I'm a little self-conscious. I'm gonna get it next week. But that made me a little doubt our exercise plan for it. That made me think, well how high am I gonna go? Someone who does already 30 minutes of rowing a day am I gonna tell them do it at like 40, 50?" So that's why I'm kind of decided to you know discuss that with you guys. And if you guys think I should keep going the workout I can but I think she's already doing what she can on the workout front.

PI1: Right, yeah. Like I said, that wasn't really reflected in the, eeh, in the input data. Like she didn't really input that as her exercise. She just like, it was like a, like, you know, point. But anyway if you like positivity is what, uh, she's going for, that makes a lot of sense.

G2: Yeah I agree. Let's change [Participant 2] to positive. Thank you guys.

The guide shares an account of the other relevant factors beyond exercise, the PWP factors which appeared highest in the top factor ranked list (albeit with a high MAPE). The guide suggests that the participant's anxiety, poor sleep and lack of social interaction might be relevant factors (lines 6-15). Then the guide highlights the two kinds of disciplinary knowledge involved in the discussion, suggesting a distinction between "stuff that I got from the interview" and the information that would be required for a more "data-driven" approach to selecting a plan. After suggesting that continuing the exercise plan has the benefit of knowing that the participant is willing to "buy in" to this plan, the PI explains that the plan should be aimed at maintaining consistency of moderate exercise.

PI1 then provides an explanation for the positive relationship between mild exercise and depressed mood, the same anomaly that was present in the data for Participant 1. PI1 suggests that mild exercise is generally not helpful for mood because it leaves "time for rumination" (lines 23-25). The guide responds by explaining that a decrease in mild exercise or a switch from mild to moderate exercise is not feasible because the participant claimed that their mild exercise comes from unavoidable day-to-day tasks such as walking from the car to the house: "they just put their daily living as their mild exercise."

Later on in the discussion (lines 62-63), the PI casts doubt on these self-reports of exercise behavior. First, the PI suggests that the cumulative daily step distance of one kilometer does not align with the notion that the participant engages only in mild exercise for "daily living." Later, in response to the guides account of the participant interview, the PI explicitly highlights this tension, saying "that wasn't really reflected in the… input data." While guides ask both Participant 1 and Participant 2 about their procedures for determining these self-report categories, the doubt with which the PI regards the self-reports of Participant 2 are unique to this case. Finally, after acknowledging the fact that the PWP guidance has technically already begun and the exercise plan has been assigned, the guide and the PI agree to switch the participant to the social connection/positivity plan.

The guide plays a key role in this interaction, which was itself an essential part of PWP plan selection and guidance. The final exchanges in this interaction (lines 62-66) make clear how the high error of the model, the lack of faith in the participant's self-reports and the guide's interjection about social connection converge toward a reassignment of the plan. In this case, the account of the guide serves as a backup mechanism for determining a plan assignment, which becomes necessary as both the statistical power of the ML system and the reliability of the

participant's self-reports come into question. While in the earlier case of Participant 1, asking follow-up questions to elicit a post-hoc kind of EMA was not viable, in this case participant reflections that were solicited by the guide become the primary basis for PWP assignment.

The PWP guide in this case demonstrates an awareness of the emphasis on empowering participants to make their own changes by transparently presenting them with their own data. The guide suggests that further analysis would be required to justify the change in plans as a "data-driven" choice, distinguishing the PWP guidance interaction from ML analysis and acknowledging the preference for the latter in the determination of the PWPs. The guide nonetheless brings about the change in plan as a result of their conversation with the participant, without which they would have continued on the exercise plan.

5. Conclusions: reflecting on two unique cases of self-reporting and PWP guidance

The above cases show how guides and PIs cope with disruptions to the trial that involve reconciling and repairing sensor data and self-reports. In cases involving both participants, one of the correlations produced by the ML model is in opposition to laboratory members' intuition—more duration of mild exercise is shown to correlate with higher depression scores. The process of managing these discrepant data takes place in two settings: PWP preparation meetings and PWP guidance sessions. The participant, their mood and their behaviors are represented differently in these two settings. In preparatory meetings, the participant appears in visualizations and spreadsheets that transparently show the basis of a recommendation without the need for complex human interpretation, even when one is offered. In guidance sessions, on the other hand, participants engage in open-ended dialog about their experience and attempts to report on these experiences. The PWP guide mediates between these settings and between the participants on one hand, who reflect on their self-reports and the experiences that they represent

in contrasting ways, and the PIs on the other, who encourage a prosaic, "data-driven" approach, using data visualizations to inform participants about their moods and behaviors.

5.1. PWP guides mediate across interactional contexts

In PWP guidance preparation, PI1 describes the participant using data visualizations that justify the choice of a particular PWP. In PWP guidance sessions, guides ask participants about their experience in a dialogic interaction. In the first of these settings, PI1 indexes various ML outputs and visualizations, instructing the guide on which are most useful for explaining the PWP selection to the participant. The "human interpretation" that PI2 offers to explain the seemingly counterintuitive relationship between mood and mild exercise in the case of Participant 1, is not recommended for sharing with the participant. Instead, PI1 suggests focusing on visualizations that do not draw attention to the sign which indicates whether a behavior is correlated with higher or lower depression score. PWP guidance preparation meetings generally accord with the formal account of the trial explored in chapters 2 & 3, according to which the intervention depends primarily on ML outputs and their capacity to activate the intrinsic motivation and inner knowledge of participants.

In PWP guidance sessions, guides take a different approach to understanding counterintuitive self-report data, asking participants for their own accounts of how they produce these reports. Participant 1 offers a rich reflection on the experience of gardening and attending exercise classes, but it leads to a communicative disjunction as the guide solicits mood information in the form of a momentary snapshot, following the logic of the EMA. Participant 2, on the other hand, offers an account of how they categorize exercise experiences that includes biomedical terminology and quantitative constructs around rate of exercise and muscle group activation. While these PWP guidance sessions consist of participant descriptions of their own experience as they are solicited in a dialog, which, as Participant 2's guide points out, is in

contrast to the "data-driven" approach of the preparatory meetings, they nonetheless end up influencing the PWP plan selected for Participant 2, as the guide interjects during the preparatory meeting to offer an account of how social interaction factors and anxiety relate to exercise for Participant 2.

These two cases demonstrate the range of explanatory interactions that can arise when participants are asked about how they report their mood and exercise. Further, both cases demonstrate the challenge of asking participants to elucidate their self-reports when the terms and logics through which they communicate about mood and exercise are not compatible with those of laboratory members and PIs. Guidance meetings offer a valuable opportunity to elicit participant feedback that could be used to improve study design and other aspects of the intervention. However, simply asking participants to clarify the way they determine their mood and exercise reports is shown to be inadequate in the case of Participant 1.

Self-tracking enthusiasts sometimes take on a conscious orientation toward this task of accurately representing activities through data, as Kristensen and Ruckenstein show in their study of self-tracking users: "Jannik remarks on the usefulness of self-tracking for a certain period of time in order to 'calibrate knowledge about the self.' As he puts it, 'By using data, you can adjust your intentions and eradicate some prejudices about yourself.' " (2018, p. 3635). This calibration process, which is challenging for individuals who seek to bring themselves into alignment with wearable technologies and in doing so, learn about and improve the self, becomes increasingly fraught as the digital health intervention recruits more and more people and technologies like, in our case, PWP guides and the PIs that train them. The present chapter demonstrates how these two kinds of data are reconciled *in interaction*, rather than through the calibration of a participant's inner faculties.

In each of these cases the PWP guide plays an important mediating role, as they are present in both laboratory meetings and PWP guidance sessions. As the link between PIs and participants, these workers take on the challenge of simultaneously acknowledging the "data-driven" emphases of the PIs in preparatory meetings, while also engaging in dialog with participants in PWP sessions themselves. Guidance work of coordinating and reconciling data system outputs and participant reflections constitutes a new kind of human labor that is required to implement ML-based interventions for mental health. This kind of mediation regarding the calibration of different kinds of and approaches to data has elsewhere been termed "data surrogacy" (Sivakumar, 2023). While this concept is applied in the context of spatial politics, PWP guidance practices exemplifies how human health workers can approach the challenge of communicating heterogenous pieces of information about behavior, mood and mental health to stakeholders with a range of frameworks for understanding these concepts.

This mediating role also involves coordinating the data-driven insights on the ML outputs alongside participant reflections on their own feelings, sensations and affects. This kind of labor exceeds the typical kinds of "invisible labor" taxonomized by Oudshoorn in the article cited above. In an article drawing attention to this kind of sensory, affective care work that is often involved in digital health projects, Maslen develops the term "sensory work." Soliciting reflections on sensations is another overlap between digital health guides and coaches and ethnographers, one camp of whom are particularly interested in developing their methods to accommodate the sensory and affective (Pink, 2015).

5.2 Data visualizations over theoretical explanation and connections to Chapter 4

Some controversial commentaries on machine learning and big data speculate on the "end of theory"—machine learning and big data driven technologies are said to provide such a complete account of the objects involved in human work that there is no longer a need to offer

general explanations. In the case of Participant 1, and the data relating to this participant, the two PI's difference in approach could be explained according to this "end of theory" hypothesis (See eg. Anderson, 2008). In broader context, however, the present case demonstrates the inevitable role of communication, interaction and interpretation in the digital psychiatry intervention, and the impossibility of a completely transparent dataset. It shifts the question away from the supposed end of theory, toward analyzing the other kinds of mediation and interaction that are recruited when scientists choose to forward visualizations of ML outputs for a participant rather than explain their cause or origin[17].

When the ML system indicates the counter-intuitive finding that mild exercise is correlated with poor mood, PI2 attempts to explain it using a theory of what caused this anomaly: perhaps the participant exercises either a mild or moderate amount each day and because the mild days lack the comparatively more beneficial moderate exercise, they are shown to cause poor mood. While we might expect the role of the guide to include sharing PI2's theory, which PI2 claims exemplifies role of human interpretation in ML implementation, the guide is instead directed toward sharing some figures over others, facilitating a two-fold mediation involving the visualizations and the PWP guidance session. Both the visualizations themselves, and the PWP guidance which is intended to transparently forward these visualizations constitute the forms of interaction, communication and interpretation that are required even as formal accounts downplay these factors in favor of simply following the data itself.

Ultimately it falls on the guide to determine how ML outputs are described, explained and interpreted. Guides are told that the theoretical explanation shared by PI2 is a valuable

[17] This approach differs from what Natasha Schüll observes of wearable technology designers who, in the absence of guides and clinicians, aspire to design technologies that affect user behavior directly: "Instead of conveying numerical data, or even patterns and correlations across different streams of data, it was essential for designers of wearable technology to 'reflect or refract data in a way that the individual can act on it' " (2016, p. 8).

contribution to the data science discussion, but not a valid kind of discourse to bring into the guidance session for which they are preparing. This finding has implications for the project of machine learning explanation and interpretation, which has been proposed as a potential safe-guard against ethics problems in data-driven medicine (Chandler et al., 2019). Selective use of particular data visualizations can take the place of theoretical *explanation*. While explanation and interpretation are often suggested as important components to add to data-driven medical interventions, in this case the human interpretation is deemed of scientific interest, but not of use to the participant, and instead the communicative and mediatory processes are left to the visualizations themselves, which could be described as expressing a kind of theory, and the PWP guides.

Not only the guidance preparation sessions, but also the guide training process that precedes them (discussed in Chapter 4), serve to determine for PWP guides what kinds of information are appropriate for the laboratory meeting context, and which are appropriate for the guidance meeting context. Chapter 4 focuses on the conversational resources used to orient guidance training toward the task of responding to participant queries as a communicative task isolated from the worldly contingencies of plan implementation. The above excerpts demonstrate the extent to which, in practice, PWP guidance exceeds the task it appeared to be in training meetings.

The importance of human interactions in the above cases also complicates common ideas about how users and technologies take on agency and power. Participants, in the above cases, are shown to be neither a powerless victim of neoliberal forces, coerced by technologies and clinicians, nor an empowered master of their self, body and health. The central role of guides and data visualizations in this trial demonstrates how the question of coercion, agency and power

often exceeds exchanges between individuals and technologies, involving new, nontraditional forms of labor, and complex discrepancies in how participants, guides and PIs conceptualize moods and activities, which reflect the dynamic interplay of bodies and particular environments during and after exercise (Kristensen & Ruckenstein, 2018; Oudshoorn, 2008).

Conclusion

The present dissertation is based on considerations of both situated practices as I observe them, and the concepts and theories that lab members use to describe their own work. While the former constitutes a rich empirical basis for characterizing what goes on at the lab and how it is practically accomplished, the latter nonetheless influences the course of this action. This approach is based on the notion that, while formal accounts become real in different ways in the messiness of particular interactional contexts, descriptions of the theory behind the trial that we hear, for example, at public talks, are importantly structured around coherent concepts. In laboratory interactions like PWP guidance trainings, on the other hand, we see how the subjects of these theories look different as they are constituted through the pragmatics of voicing and footing. The findings of this dissertation are the result of exploring the gaps in coherent theories that lab members use to describe their work, looking at science not as a discourse that variously represents the things of the world, but rather a practice that takes place within particular environments.

The conceptualization of the trial that we hear as the PIs describe it centers around the joint capacities of a novel ML-based system and the participants that it empowers. In Chapter 2, this formal account is explored as a combination of what sociologists of health and medicine call healthism, which emphasizes the responsibility of the individual patient, and techno-solutionism, which promises powerful outcomes as a result of new technologies themselves. PIs reconcile

discourses that are characterized by these two terms from social theory in their accounts of the trial as empowering participants through the digital system.

In Chapter 3, the participant side of this two-sided formulation is investigated further. The participants of the TPL trial are described as containing within them the capacity to understand relationships between behavior and mood, the motivation to make changes on the basis of these relationships, and also the experience of positive mood itself. By expanding the characterization of the participant into a multi-layered or multiple entity, the laboratory can explain the complex process of the intervention largely in terms of the accomplishments of individual participants. Building on the historical context from Chapter 1, this notion of a multi-layered participant is shown to be another version of a representational theory of multiple meanings, and is put into the context of a broader shift in psychiatry whereby multiple meanings of patients' inner selves have been replaced by multiple versions or layers.

Philosophical pragmatism is used to show how these multiples arise in the face of complexity, and to suggest a way out that is empirically grounded. Multiple versions of an entity can be evoked to cope as entities change in relation to other people and things, and are measured in different ways in different contexts. An alternative is suggested in ethnographic approaches which aim to specify entities in particular contexts in which they are put into relation with others.

Observational approaches that specify situated scientific practices are generally taken up to shed light on otherwise overlooked aspects of the intervention, including the important work of PWP guides, whose work is largely described as technical, as I discuss in Chapter 2. Also important to the success of the intervention, but largely unacknowledged as lab members describe the trial, are the environments, neighborhoods and material infrastructures in which participants implement the intervention, which are described in Chapters 3 and 5.

The technical or mediatory nature of PWP guides that we hear in laboratory descriptions is a partial account, and yet it is also manifest in the interactional reality of guidance training meetings, in which PIs use conversational resources like hypothetical active voicing (HAV) to orient discussion around the task of appropriately responding to participant queries, even when their questions venture into issues of study design and broad implementation issues. In sum, Chapter 4 addresses the question of how PIs practically accomplish the relegation of the guidance role to a narrowly understood task of saying the right thing to participants.

In guidance meetings themselves, however, guides exceed the formal account of their work and the manner in which they have been trained. Chapter 5 demonstrates not only the presence and engaged listening of guides, but also their key role in mediating between PIs and participants. Guides are shown to attempt to solicit post-hoc EMA-style reflections from participants to address breakages and inconsistencies in ML outputs. Also in this chapter, participants are shown to offer differing approaches to the task of justifying their self-reports of mood and exercise, with one avoiding the logic of the EMA and another making use of biomedical and quantitative language to become intelligible to the guide as a part of the data-driven sociotechnical system. Finally, Chapter 5 demonstrates how data visualizations can be used to guide conversations with participants, in this case away from discussion of the "sign" or direction of a correlation between behavior and mood.

These findings offer insight on pressing questions in the social science of digital health, including the topic of explanation and interpretation in AI-driven health, the question of automation and new labor, and the implementational reality of implementing digital health interventions in real world environments.

Having explored the history of meaning and explanation in the psychiatric discipline, the problem of better explaining and interpreting ML outputs in digital mental health is made to seem like a fundamental rather than incidental disjunction. The TPL trial is shown to be the continuation of a trajectory that goes back to the decline of psychoanalysis in which questions of meaning and interpretation have undergone a devaluation in favor of a more prosaic data-driven approach. Rather than reverting to old psychiatric concepts around the paradoxes of linguistic meaning that underlie psychiatric disorders, the way out of this challenge is through robust accounts of the experiences through which participants implement ML-based psychiatry interventions in their living and working places, the ways they calibrate their senses of mood and behavior to accord with digital systems and the interactions with other people.

The individualized nature of the intervention makes the data producing efforts of participants challenging the observe. However, the present study demonstrates how interactions in which participants discuss the trial as they attempt to implement it, invite an unpredictable kind of context. Unstructured, interpretive discussions around behavior and mood, like those that sometimes happen in PWP guidance meetings, lead participants toward new understandings of the context in which the trial takes place in each particular case. The value of these interactions for the present study comes not from their potential to reveal in the participant's psyche a hidden etiology, drive or schema, but rather the potential to better understand the trial as a practical enterprise that is carried out in real world environments.

Related to this notion is the increasingly prominent movement for explainability in AI. These efforts are often directed toward objectives questions like, for example, explaining why someone was denied a loan by a bank's algorithmic decision system. Some scholars have pointed to a more practical notion of explanation that is grounded in the power relations of a particular

site. Scott Robbins uses the example of a visual explanation of a cancer diagnosis which is, in the case of its presentation to a patient, a useless attempt to reconfigure human-machine control, given the broader power structure of the medical model: "A medical diagnosis algorithm that classifies someone as having a brain tumor might, for example, provide a heat map of which parts of the brain scan most contributed to the diagnosis. This 'explanation' would probably be useless to a patient—or to anyone else without very specific medical training. However, if the goal is that the algorithm is under 'meaningful human control' then we are not concerned with the patient's understanding of the explanation" (Robbins, 2019, p. 502). For Robbins, the explanation must be offered to the human actor who can use it to effectively control the non-human machinery.

In another article aimed at expanding the notion of explainability to take into account social context, Ehsan and colleagues interviewed AI developers and users to get a sense of how they value explanations as they use AI. From these interviews emerged a common interest in the 4 W's: why, who, what, and when. The authors explain the importance of answering these questions using what they call a "social transparency" model, which is based on giving users access to transparent representations of their own identities, the content they access and the interactions that are facilitated by the AI platform. By combining the "why" questions at the heart of explanability theory with "who," what," and "when," the authors aim to offer explanations that are grounded in the sociotechnical arrangements of the particular AI-based intervention. Timestamps on "comments" that explain AI decision, for example, answer the "when" question: "Knowing the when 'puts things into perspective' (P16-NS) because it adds 'context to the decision and strengthen[s] the why' (P18-S)" (Ehsan et al., 2021, p. 13).

What is sought in these explanation systems is different from the value we see in the interactions between PWP guides and participants, which invite discussion of the environments in which participants enact the trial, and the interactions that constitute it. Robbins' work usefully draws attention to the political dynamics of medicine, by which patients cannot be simply empowered by accessing technical visualizations of medical data. However, the kind of discussion of ML we see in PWP guidance cannot be adequately described in terms of *which actors have the power* to alter the course of the intervention, as the these meetings themselves become sites in which these roles are continually contested and renewed.

As we saw in chapters 4 and 5, participants use PWP guidance sessions to raise issues with guides and negotiate how they will produce data regarding mood and exercise. When guides interact with the PIs we again see terrain for continual contestation, where PIs use particular conversational resources to establish and maintain the role of the PWP guide as one based on resolving participant queries. Broadly, answers to the question of relevant actors are shown to differ greatly as we compare formal accounts from the lab and observed practices—chapters 2 and 3 show how the participant who is empowered by the data system is a central actor as the trail is described, whereas PWP guides are shown to play a constitutive role as the trial is carried out. In this case, efforts to improve explanability cannot be oriented toward a ready-made structural understanding of power relations as these relations only become real in the course of the interactions that constitute the intervention itself.

The "4 Ws" approach of Ehsan and colleagues faces a similar limitation. Pairing comments that explain a recommendation with a timestamp, to take "when" as an example, does not offer a contextual understanding of how the sociotechnical system developed a recommendation *over* time. As we see in chapter 5, participants express a range of complex

understandings of how mood and behavior relate temporally as they attempt to justify their use of the ecological momentary assessment, drawing guides' attention to the problematic temporal relationship between embodied experience and "snapshots" of mood and exercise. Further, as we saw in chapter 4, the way that roles are established in the lab depends on the unfolding of particular conversational strategies which include various footings, placing interactional figures in an imagined future guidance scenario, for example.

Environmental contingencies and relationships between people come to the fore when we attend to the interactions by which the trial is carried out. This is a result of the inevitably indexical nature of naturally occurring interaction (Cicourel, 1974; Silverstein, 2003). Rather than developing questions or fields regarding these factors to be "filled in" as the intervention gets underway, human interactions should be taken as valuable components of interventions, which should be structured to invite the wide-ranging kinds of reflections we saw in PWP guidance meetings. While these meetings offered an opening for this productive kind of interaction, they were not encouraged as such. While we could consider the formal accounts of the PIs as just a limiting conceptual framework, instead we will look to see the kinds of interactional moves that give this formal concept an empirical presence in the trial. In chapter 4 we see how HAV and other conversational resources become tools for establishing and maintaining guidance as a query-resolving role that is supplemental to the technical system. It is though this interactional process that the general notion that "PWP guidance isn't therapy" becomes empirically observable.

However, in observation of PWP guidance sessions, this guidance labor is shown to be more than a technical or mediatory component. Rigorously specifying the work of these kinds of nontraditional health professionals is important as health systems that are promoted as based on

AI and automation often recruit new kinds of labor that goes unacknowledged or underspecified. The present dissertation provides a framework for considering the formal accounts through which this kind of interactional work is minimized, as well as the kind of observation and engagement required to account for it in its complexity. The practice-oriented account demonstrates how, what is sometimes considered a broader sociological phenomenon of the minimization and relegation of nontraditional forms of human labor, becomes real in particular interactions.

One key difference between accounts of guidance labor as supplementary and technical and the way it appears in observation, is the issue of environmental contexts and engagement with worldly environments. PWP guidance is shown to be a site for taking into account the contingencies that emerge as participants use the digital mental health system in particular environments. The kinds of neighborhoods, climates, working hours and other endless contingencies of different living and working environments become relevant only in the open-ended dialogs of PWP guidance. Observing these sessions offers a chance to complicate the notion of a maximally portable digital health technology, which is promoted for its neutrality with regard to the context in which it is implemented. While these meetings can, as shown above in Chapter 5, have an influence on the course of the PWP recommendation, the reflections of participants could be integrated into the study design to better understand implementational problems that participants experience as more than questions or complaints that need to be responded to.

Because of the remote work approach of the laboratory and its extensive use of digital tools, studying it has yielded methodological insights on digital, remote and virtual observation, to the extent that this is possible as such. In Chapter 2, for instance, we see the limitations of

gesture and body movements on Zoom teleconference, as the two PIs who share a camera exchange glances in a way other interlocutors cannot, thus embodying a disciplinary disjunction in how the laboratory balances the agencies of participants the ML system. Similarly, in Chapter 5 we see how screen sharing abilities play a major role in the course of interaction, as a PI demonstrates to guides which figures they should show participants, moving from image to image in a single turn of conversation. This interactional context is not just the result of my observations being conducted remotely, but the parallel situation of the scientists I study—the challenges of following gestures and body movements and the power of screen sharing capabilities are both a new part of my research approach and that of the lab members at the TPL. This parallel exemplifies Garfinkel's notion of unique adequacy. For the competent participant, activities always include their own technologies for involvement and observation.

In a broader methodological connection, my use of observational approaches in the lab parallels, in some ways, other laboratory practices that I aim to specify, as both my observations and PWP guidance feature engagement with participants in open-ended interaction. The ways these activities are left out of laboratory descriptions or made to seem technical has implications for the prospect of integrating qualitative, observational approaches into projects that are otherwise described as computational (Beaulieu, 2017; Bemme et al., 2020). The appropriateness of both an observational and also an ethnomethodological perspective on this trial is not a precondition to be established in advance of the study, but rather a fact that can only be evidenced by the empirical material of the study itself. Only by attending to interactions as they unfold can we demonstrate that the trial is constituted and made observable through these interactions and therefore, that these interactions are worth studying.

Further work is needed on topics relating to the ML-based digital mental health trial. The way that participants use the wearable device and the smart phone app could be more rigorously specified by observing these practices as they occur in everyday environments. This would give a better perspective on the user experience of the application, the embodied experience of wearable device use and the challenges of satisfying the various data reporting requirements. These insights would strengthen our account of PWP guidance interactions by giving another sense of the experiences that participants attempt to relate to PWP guides in these interactions. This could also be achieved autoethnographically or using the "walk-through" method which was designed to give a sense of the emplaced experience of digital application use (Light et al., 2018). While the present approach effectively demonstrates the problems that arise as participants and lab members try to develop cohesive accounts of how participants use their devices and complete EMAs (see eg. Chapter 5), it would be enhanced by the inclusion of empirical material from the site of everyday data production.

Another area for further investigation is the use of teleconferencing tools in lab meetings and PWP guidance. While this is briefly touched on in chapter 2, as we see the PIs exchange glances in a way afforded by their shared camera, it could be further explored using a cognitive semiotic approach that could specify the blending of the shared virtual space of the teleconference screen and the homes, offices, labs and classrooms in which interlocutors are physically located (Hougaard et al., 2022).

Also of interest for further work are studies that attempt digital mental health interventions on larger scales, including those that use a control arm and those that are integrated into public and municipal projects. The use of ML-based technology for larger trials would be of interest because of the greater relevance of constructs around clinical validity. These are less

important in the present trial because of its exploratory nature. Some mobile apps for mental health are being deployed through community organizations and governments—this is another relevant context for studying this kind of intervention, as it brings with it novel challenges regarding the specification of new forms of human labor (Gratzer et al., 2021).

Another promising area for further work on this topic is around linguistic differences and differences in language background and the role these factors play in the interactional course of implementing digital mental health interventions. This work could form part of a larger study that recruits participants for diversity of cultural and linguistic background. This kind of study, rather than focusing on language ideology broadly constructed, could investigate the course of interaction as multi-lingual participants, guides and scientists describe data, behavior and mood in English and other languages.

References

Agha, A. (2005). Voice, Footing, Enregisterment. *Journal of Linguistic Anthropology*, *15*(1), 38–59. https://doi.org/10.1525/jlin.2005.15.1.38

Ajana, B. (2017). Digital health and the biopolitics of the Quantified Self. *DIGITAL HEALTH*, *3*, 2055207616689509. https://doi.org/10.1177/2055207616689509

Alac, M. (2011). *Handling Digital Brains*. MIT Press.

Alač, M. (2020). Beyond intersubjectivity in olfactory psychophysics I: Troubles with the Subject. *Social Studies of Science*, *50*(3), 440–473. https://doi.org/10.1177/0306312720915645

Albee, G. W. (2000). The Boulder model's fatal flaw. *American Psychologist*, *55*(2), 247–248. https://doi.org/10.1037/0003-066X.55.2.247

Alfaras, M., Tsaknaki, V., Sanches, P., Windlin, C., Umair, M., Sas, C., & Höök, K. (2020). From Biodata to Somadata. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3313831.3376684

Algera, E. (2023). Knowing (with) the body: Sensory knowing in contraceptive self-tracking. *Sociology of Health & Illness*, *45*(2), 242–258. https://doi.org/10.1111/1467-9566.13570

Ameka, F. (1992). Interjections: The universal yet neglected part of speech. *Journal of Pragmatics*, *18*(2–3), 101–118.

Anderson, C. (2008). The End of Theory: The data Deluge Makes the Scientific Method Obsolete. *Wired*.

Andrejevic, M. (2013). Alienation's returns. In C. Fuchs & M. Sandoval (Eds.), *Critique, Social Media and the Information Society* (pp. 179–190). Routledge.

Austin, J. L. (1975). *How to do things with words* (Vol. 88). Oxford university press.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

Banner, O. (2019). Technopsyence and Afro-Surrealism's Cripistemologies. *Catalyst: Feminism, Theory, Technoscience*, *5*(1), 1–29. https://doi.org/10.28968/cftt.v5i1.29612

Baños, R. M., Herrero, R., & Vara, M. D. (2022). What is the Current and Future Status of Digital Mental Health Interventions? *The Spanish Journal of Psychology*, *25*, e5. https://doi.org/10.1017/SJP.2022.2

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Barnes, R., & Moss, D. (2007). Communicating a feeling: The social organization of `private thoughts'. *Discourse Studies*, *9*(2), 123–148. https://doi.org/10.1177/1461445607075339

Baudrillard, J. (2008). *The Ecstasy of Communication* (S. Lotringer, Ed.; Repr.). Semiotext(e).

Baumgartner, R. (2021). Precision medicine and digital phenotyping: Digital medicine's way from more data to better health. *Big Data & Society*, *8*(2), 20539517211066452. https://doi.org/10.1177/20539517211066452

Beaulieu, A. (2017). Vectors for Fieldwork: Computational Thinking and New Modes of Ethnography. In H. Horst, L. Hjorth, A. Galloway, & B. Genevieve (Eds.), *The Routledge Companion to Digital Ethnography* (pp. 29–39). Routledge.

Beck, A. T. (1963). Thinking and Depression: I. Idiosyncratic Content and Cognitive Distortions. *Archives of General Psychiatry*, *9*(4), 324. https://doi.org/10.1001/archpsyc.1963.01720160014002

Beck, A. T. (1964). Thinking and Depression: II. Theory and Therapy. *Archives of General Psychiatry*, *10*(6), 561. https://doi.org/10.1001/archpsyc.1964.01720240015003

Beck, J. S. (2021). *Cognitive behavior therapy: Basics and beyond* (Third edition). The Guilford Press.

Beck, U., Giddens, A., & Lash, S. (1994). *Reflexive modernization: Politics, tradition and aesthetics in the modern social order*. Stanford University Press.

Becvar, L. A., Hollan, J., & Hutchins, E. (2005). Hands as molecules: Representational gestures used for developing theory in a scientific laboratory. *Semiotica*, *2005*(156). https://doi.org/10.1515/semi.2005.2005.156.89

Bemme, D., Brenman, N., & Semel, B. (2020, October 13). The subjects of digital psychiatry. *Somatosphere*. http://somatosphere.net/2020/subjects-of-digital-psychiatry.html/

Berisha, V., Krantsevich, C., Hahn, P. R., Hahn, S., Dasarathy, G., Turaga, P., & Liss, J. (2021). Digital medicine and the curse of dimensionality. *Npj Digital Medicine*, *4*(1), Article 1. https://doi.org/10.1038/s41746-021-00521-5

Berkhout, S., & Zaheer, J. (2021). Digital Self-Monitoring, Bodied Realities: Re-Casting App-Based Technologies in First Episode Psychosis. *Catalyst: Feminism, Theory, Technoscience*, *7*(1). https://doi.org/10.28968/cftt.v7i1.34101

Birk, R., & Samuel, G. (2020). Can digital data diagnose mental health problems? A sociological exploration of 'digital phenotyping.' *Sociology of Health & Illness*, *42*(8), 1873–1887. https://doi.org/10.1111/1467-9566.13175

Bossen, C., Pine, K. H., Cabitza, F., Ellingsen, G., & Piras, E. M. (2019). Data work in healthcare: An Introduction. *Health Informatics Journal*, *25*(3), 465–474. https://doi.org/10.1177/1460458219864730

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. MIT press.

Brown, N., & Michael, M. (2003). A Sociology of Expectations: *Retrospecting Prospects and Prospecting Retrospects*. *Technology Analysis & Strategic Management*, *15*(1), 3–18. https://doi.org/10.1080/0953732032000046024

Bruner, J., Goodnow, J., & Austin, G. (1956). *A Study of Thinking*. John Wiley & Sons, Inc.

Bruner, J. S. (1990). *Acts of meaning*. Harvard University Press.

Callon, M., & Rabeharisoa, V. (2008). The Growing Engagement of Emergent Concerned Groups in Political and Economic Life: Lessons from the French Association of Neuromuscular Disease Patients. *Science, Technology, & Human Values*, *33*(2), 230–261. https://doi.org/10.1177/0162243907311264

Chandler, C., Foltz, P. W., & Elvevåg, B. (2019). Using Machine Learning in Psychiatry: The Need to Establish a Framework That Nurtures Trustworthiness. *Schizophrenia Bulletin*, sbz105. https://doi.org/10.1093/schbul/sbz105

Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021). Ethical Machine Learning in Health Care. *Annual Review of Biomedical Data Science*, *4*(1), 123–144. https://doi.org/10.1146/annurev-biodatasci-092820-114757

Cicourel, A. V. (1974). *Cognitive Sociology: Language and Meaning in Social Interaction*. The Free Press.

Clark, D. A., & Beck, A. T. (1999). *Scientific foundations of cognitive theory and therapy of depression*. John Wiley.

Clayman, S., & Maynard, D. (1995). Ethnomethodology and Conversation Analysis. In *Situated order: Studies in the social organization of talk and embodied activities* (pp. 1–30).

Cloninger, S. C. (1996). *Personality: Description, dynamics, and development*. W.H. Freeman and Company.

Cohen-Cole, J. (2005). The reflexivity of cognitive science: The scientist as model of human nature. *History of the Human Sciences*, *18*(4), 107–139. https://doi.org/10.1177/0952695105058473

Collins, R. (1981). Micro-translation as a theory-building strategy. In *Advances in social theory and methodology: Toward an integration of micro- and macro-sociologies* (pp. 81–108). Routledge & Kegan Paul.

Crawford, R. (1980). Healthism and the Medicalization of Everyday Life. *International Journal of Health Services*, *10*(3), 365–388. https://doi.org/10.2190/3H2H-3XJN-3KAY-G9NY

Deleuze, G., & Guattari, F. (1983). *Anti-Oedipus: Capitalism and schizophrenia*. University of Minnesota Press.

Della Bianca, L. (2021). The Cyclic Self: Menstrual Cycle Tracking as Body Politics. *Catalyst: Feminism, Theory, Technoscience*, *7*(1). https://doi.org/10.28968/cftt.v7i1.34356

Dobson, K., & Dozois, D. (2010). Historical and philosophical bases of the Cognitive-behavioral therapies. In *Handbook of cognitive-behavioral therapies* (3rd edition, pp. 3–38). Guilford Press.

Dudhwala, F., & Woolgar, S. (2019). Responsibility, response-ability and responsivity: The new characteristics of accountability in the face of online patient feedback – ethnographic case studies in four NHS trusts. In J. Powell, H. Atherton, V. Williams, F. Mazanderani, A.-M. Boylan, J. Fleming, S. Kirkpatrick, A. Martin, M. van Velthoven, A. de Iongh, D. Findlay, L. Locock, & S. Ziebland (Eds.), *Using online patient feedback to improve NHS services: The INQUIRE multimethod study*. NIHR Journals Library. https://www.ncbi.nlm.nih.gov/books/NBK549403/

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding Explainability: Towards Social Transparency in AI systems. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–19. https://doi.org/10.1145/3411764.3445188

Eisenberg, L. (1986). Mindlessness and Brainlessness in Psychiatry. *British Journal of Psychiatry*, *148*(5), 497–508. https://doi.org/10.1192/bjp.148.5.497

Ellis, A. (1963). *Reason and emotion in psychotherapy*. The Institute for Rational Living Inc.

Ericson, R. V., & Haggerty, K. D. (2000). The surveillant assemblage. *British Journal of Sociology*, *51*(4), 605–622.

Fink, B. (2007). *Fundamentals of psychoanalytic technique: A Lacanian approach for practitioners* (1st ed). W.W. Norton.

Foucault, M. (2008). *The birth of biopolitics: Lectures at the Collège de France, 1978-79* (M. Senellart, Ed.). Palgrave Macmillan.

Freud, S. (1999). *The standard edition of the complete psychological works of Sigmund Freud* (J. Strachey, Ed.; Repr, Vol. 12). Hogarth Press.

Friedman, R. J., & Katz, M. M. (Eds.). (1974). *The Psychology of depression: Contemporary theory and research*. Winston; [distributed by Halsted Press Division, Wiley].

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*(10), 474–480. https://doi.org/10.1016/j.tics.2005.08.011

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Prentice-Hall.

Garfinkel, H., & Wieder, L. (1992). Two Incommensurable, Asymmetrically Alternate Technologies of Social Analysis. In *Text in Context: Contributions to Ethnomethodology* (pp. 175–206). Sage Publications.

Geiger, R. S., & Ribes, D. (2011). Trace Ethnography: Following Coordination through Documentary Practices. *2011 44th Hawaii International Conference on System Sciences*, 1–10. https://doi.org/10.1109/HICSS.2011.455

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

Goodwin, C. (1979). The interactional construction of a sentence in natural conversation. In *Everyday Language: Studies in Ethnomethodology* (pp. 97–121).

Goodwin, C. (1994). Professional Vision. *American Anthropologist*, *96*(3), 606–633.

Goodwin, C. (2003). Embedded Context. *Research on Language & Social Interaction*, *36*(4), 323–350. https://doi.org/10.1207/S15327973RLSI3604_2

Gratzer, D., Torous, J., Lam, R. W., Patten, S. B., Kutcher, S., Chan, S., Vigo, D., Pajer, K., & Yatham, L. N. (2021). Our Digital Moment: Innovations and Opportunities in Digital Mental Health Care. *The Canadian Journal of Psychiatry*, *66*(1), 5–8. https://doi.org/10.1177/0706743720937833

Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.

Green, S., Hillersdal, L., Holt, J., Hoeyer, K., & Wadmann, S. (2022). The practical ethics of repurposing health data: How to acknowledge invisible data work and the need for prioritization. *Medicine, Health Care and Philosophy*. https://doi.org/10.1007/s11019-022-10128-6

Grosjean, S., Matte, F., & Nahon-Serfaty, I. (2021). The Two-Sides of Video-Ethnography for Studying "Sensing-at-Distance." In *Organizational Video-Ethnography Revisited: Making Visible Material, Embodied and Sensory Practices* (pp. 59–75). Springer International Publishing. https://doi.org/10.1007/978-3-030-65551-8

Hacking, I. (2002). *Historical ontology*. Harvard University Press.

Harris, G. (2011, March 5). Talk Doesn't Pay, So Psychiatry Turns Instead to Drug Therapy. *The New York Times*.

Hartung, G. (Ed.). (2015). *From Hegel to Windelband: Historiography of philosophy in the 19th century*. De Gruyter.

Heath, C., & Hindmarsh, J. (2002). Analysing Interaction: Video, Ethnography and Situated Conduct. In T. May (Ed.), *Qualitative Research in Action* (pp. 100–121). SAGE Publications Ltd. https://doi.org/10.4135/9781849209656.n4

Hegel, G. W. F. (2013). *Phenomenology of spirit* (A. V. Miller, Trans.; Reprint.). Oxford Univ. Press.

Hirsch, T., Merced, K., Narayanan, S., Imel, Z. E., & Atkins, D. C. (2017). Designing Contestability: Interaction Design, Machine Learning, and Mental Health. *Proceedings of the 2017 Conference on Designing Interactive Systems*, 95–99. https://doi.org/10.1145/3064663.3064703

Holt, E. (1996). Reporting on Talk: The Use of Direct Reported Speech in Conversation. *Research on Language & Social Interaction*, *29*(3), 219–245. https://doi.org/10.1207/s15327973rlsi2903_2

Hougaard, A., Oakley, T., & Coulson, S. (2022). Who's talking? Cognitive semiotics in the (new media) wild. *Cognitive Semiotics*, *15*(1), 47–73. https://doi.org/10.1515/cogsem-2022-2003

Hurlburt, R. T., & Knapp, T. J. (2006). Münsterberg in 1898, Not Allport in 1937, Introduced the Terms 'Idiographic' and 'Nomothetic' to American Psychology. *Theory & Psychology*, *16*(2), 287–293. https://doi.org/10.1177/0959354306062541

Insel, T. R. (2017). Digital Phenotyping: Technology for a New Science of Behavior. *JAMA*, *318*(13), 1215–1216. https://doi.org/10.1001/jama.2017.11295

Irani, L. C., & Silberman, M. S. (2013). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 611–620. https://doi.org/10.1145/2470654.2470742

Jain, S. H., Powers, B. W., Hawkins, J. B., & Brownstein, J. S. (2015). The digital phenotype. *Nature Biotechnology*, *33*(5), Article 5. https://doi.org/10.1038/nbt.3223

Jefferson, G. (2004). Glossary of transcript symbols. *Conversation Analysis: Studies from the First Generation*, 24–31.

Kahlert, D. (2016). Built Environment, Physical Activity and Social Participation of Older People. In S. Selke (Ed.), *Lifelogging* (pp. 189–204). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-13137-1_10

Karbelnig, A. M. (2022). Chasing infinity: Why clinical psychoanalysis' future lies in pluralism. *The International Journal of Psychoanalysis*, *103*(1), 5–25. https://doi.org/10.1080/00207578.2021.1975288

Katz, J. (2010). Time for new urban ethnographies. *Ethnography*, *11*(1), 25–44.

Keane, P. A., & Topol, E. J. (2018). With an eye to AI and autonomous diagnosis. *Npj Digital Medicine*, *1*(1), Article 1. https://doi.org/10.1038/s41746-018-0048-y

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17*(1), 195. https://doi.org/10.1186/s12916-019-1426-2

Koskela, I., & Palukka, H. (2011). Trainer interventions as instructional strategies in air traffic control training. *Journal of Workplace Learning*, *23*(5), 293–314. https://doi.org/10.1108/13665621111141902

Koutsouleris, N., Hauser, T. U., Skvortsova, V., & De Choudhury, M. (2022). From promise to practice: Towards the realisation of AI-informed mental health care. *The Lancet Digital Health*, *4*(11), e829–e840. https://doi.org/10.1016/S2589-7500(22)00153-4

Kristensen, D. B., & Ruckenstein, M. (2018). Co-evolving with self-tracking technologies. *New Media & Society*, *20*(10), 3624–3640. https://doi.org/10.1177/1461444818755650

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed). University of Chicago Press.

Lacan, J. (1998). *The four fundamental concepts of psychoanalysis* (J.-A. Miller, Trans.). W.W. Norton & Company.

Lamiell, J. T. (1998). `Nomothetic' and `Idiographic': Contrasting Windelband's Understanding with Contemporary Usage. *Theory & Psychology*, *8*(1), 23–38. https://doi.org/10.1177/0959354398081002

Laplanche, J. (1987). Nouveaux fondements pour la psychanalyse: La séduction originaire. *PUF*.

Lave, J. (1997). The culture of acquisition and the practice of understanding. Reprinted in D. Kirshner & JA Whitson (Eds.), Situated cognition: Social, semiotic, and psychological perspectives. In D. Kirshner & J. Whitson (Eds.), *Situated cognition: Social, semiotic, and psychological perspectives*. Mahwah, NJ: Erlbaum.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Leahy, R. L. (2001). *Overcoming resistance in cognitive therapy*. Guilford Press.

Levinson, S. (2012). Foreward. In *Language, Thought and Reality: Selected Writing of Benjamin Lee Whorf* (2nd edition, pp. vii–xxiii). MIT Press.

Leyland, C. (2016). 'Pre-enactment' in team-teacher planning talk: Demonstrating a possible future in the here-and-now. *Pragmatics*, *26*, 675–704. https://doi.org/10.1075/prag.26.4.07ley

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, *20*(3), 881–900. https://doi.org/10.1177/1461444816675438

Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., & Fuller-Tyszkiewicz, M. (2019). The efficacy of app-supported smartphone interventions for mental health problems: A meta-analysis of randomized controlled trials. *World Psychiatry*, *18*(3), 325–336. https://doi.org/10.1002/wps.20673

Lupton, D. (2013a). Quantifying the body: Monitoring and measuring health in the age of mHealth technologies. *Critical Public Health*, *23*(4), 393–403. https://doi.org/10.1080/09581596.2013.794931

Lupton, D. (2013b). The digitally engaged patient: Self-monitoring and self-care in the digital health era. *Social Theory & Health*, *11*(3), 256–270. https://doi.org/10.1057/sth.2013.10

Lupton, D. (2014). Beyond Techno-Utopia: Critical Approaches to Digital Health Technologies. *Societies*, *4*(4), Article 4. https://doi.org/10.3390/soc4040706

Lupton, D., & Maslen, S. (2018). The more-than-human sensorium: Sensory engagements with digital self-tracking technologies. *The Senses and Society*, *13*(2), 190–202. https://doi.org/10.1080/17458927.2018.1480177

Lupton, D., & Watson, and A. (2021). Towards more-than-human digital data studies: Developing research-creation methods. *Qualitative Research*, *21*(4), 463–480. https://doi.org/10.1177/1468794120939235

Lussier, A. (1991). The search for common ground: A critique. *International Journal of Psycho-Analysis*, *72*, 57–62.

Lynch, M. (1993). *Scientific practice and ordinary action: Ethnomethodology and social studies of science*. Cambridge University Press.

Marent, B., & Henwood, F. (2023). Digital health: A sociomaterial approach. *Sociology of Health & Illness*, *45*(1), 37–53. https://doi.org/10.1111/1467-9566.13538

Marsh, J. A., Bertrand, M., & Huguet, A. (2015). Using Data to Alter Instructional Practice: The Mediating Role of Coaches and Professional Learning Communities. *Teachers College Record*, *117*(4), 1–40. https://doi.org/10.1177/016146811511700411

Maslen, S. (2017). Layers of sense: The sensory work of diagnostic sensemaking in digital health. *Digital Health*, *3*, 2055207617709101. https://doi.org/10.1177/2055207617709101

Maslen, S., & Harris, A. (2021). Becoming a diagnostic agent: A collated ethnography of digital-sensory work in caregiving intra-actions. *Social Science & Medicine*, *277*, 113927. https://doi.org/10.1016/j.socscimed.2021.113927

McCradden, M. D., Anderson, J. A., A. Stephenson, E., Drysdale, E., Erdman, L., Goldenberg, A., & Zlotnik Shaul, R. (2022). A Research Ethics Framework for the Clinical Translation of Healthcare Machine Learning. *The American Journal of Bioethics*, *22*(5), 8–22. https://doi.org/10.1080/15265161.2021.2013977

Mills, C., & Hilberg, E. (2020). The construction of mental health as a technological problem in India. *Critical Public Health*, *30*(1), 41–52. https://doi.org/10.1080/09581596.2018.1508823

Mol, A. (2002). *The body multiple: Ontology in medical practice*. Duke University Press.

Mol, A., Moser, I., & Pols, J. (Eds.). (2010a). *Care in practice: On tinkering in clinics, homes and farms*. Transcript.

Mol, A., Moser, I., & Pols, J. (2010b). Care: Putting practice into theory. In *Care In Practice: On Tinkering in Clinics, Homes and Farms* (pp. 7–26). Transaction Publishers.

Mort, M., May, C. R., & Williams, T. (2003). Remote Doctors and Absent Patients: Acting at a Distance in Telemedicine? *Science, Technology, & Human Values*, *28*(2), 274–295. https://doi.org/10.1177/0162243902250907

Nafus, D. (Ed.). (2016). *Quantified: Biosensing technologies in everyday life*. The MIT Press.

National Advisory Mental Health Council. (n.d.). *Opportunities and Challenges of Developing Information Technologies on Behavioral and Social Science Clinical Research*. Retrieved August 12, 2023, from https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/opportunities-and-challenges-of-developing-information-technologies-on-behavioral-and-social-science-clinical-research

National Institute for Mental Health Strategic Plan for Research (20-MH-8096; p. 54). (2022). National Institute for Mental Health. https://www.nimh.nih.gov/sites/default/files/documents/about/strategic-planning-reports/NIMH%20Strategic%20Plan%20for%20Research_2022_0.pdf

Noë, A. (2004). *Action in perception*. MIT press.

Norcross, J. C., & Newman. (1992). Psychotherapy integration: Setting the context. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psychotherapy integration* (pp. 3–45). Basic Books.

Onnela, J.-P. (2021). Opportunities and challenges in the collection and analysis of digital phenotyping data. *Neuropsychopharmacology*, *46*(1), Article 1. https://doi.org/10.1038/s41386-020-0771-3

Osbeck, L. M., & Nersessian, N. J. (2006). The Distribution of Representation. *Journal for the Theory of Social Behaviour*, *36*(2), 141–160. https://doi.org/10.1111/j.1468-5914.2006.00301.x

Oudshoorn, N. (2008). Diagnosis at a distance: The invisible work of patients and healthcare professionals in cardiac telemonitoring technology. *Sociology of Health & Illness*, *30*(2), 272–288.

Pagano, R., Cremonesi, P., Larson, M., Hidasi, B., Tikk, D., Karatzoglou, A., & Quadrana, M. (2016). The Contextual Turn: From Context-Aware to Context-Driven Recommender Systems. *Proceedings of the 10th ACM Conference on Recommender Systems*, 249–252. https://doi.org/10.1145/2959100.2959136

Paris, J. (2005). *Fall of an icon: Psychoanalysis and academic psychiatry*. University of Toronto Press.

Pedersen, T., & Johansen, C. (2020). Behavioural artificial intelligence: An agenda for systematic empirical studies of artificial inference. *AI & SOCIETY*, *35*(3), 519–532. https://doi.org/10.1007/s00146-019-00928-5

Pickersgill, M. (2011). 'Promising' therapies: Neuroscience, clinical practice, and the treatment of psychopathy. *Sociology of Health & Illness*, *33*(3), 448–464. https://doi.org/10.1111/j.1467-9566.2010.01286.x

Pickersgill, M. (2019). Digitising psychiatry? Sociotechnical expectations, performative nominalism and biomedical virtue in (digital) psychiatric praxis. *Sociology of Health & Illness*, *41*(S1), 16–30. https://doi.org/10.1111/1467-9566.12811

Pink, S. (2015). *Doing sensory ethnography*. Sage. https://books.google.com/books?hl=en&lr=&id=_sSICwAAQBAJ&oi=fnd&pg=PP1&dq=pink+sensory+ethnography&ots=wp93POF729&sig=cj4POk6BEla7MuMD_vOqTjdGLHk

Pink, S., Ruckenstein, M., Willim, R., & Duque, M. (2018). Broken data: Conceptualising data in an emerging world. *Big Data & Society*, *5*(1), 2053951717753228. https://doi.org/10.1177/2053951717753228

Pols, J., Willems, D., & Aanestad, M. (2019). Making sense with numbers. Unravelling ethico-psychological subjects in practices of self-quantification. *Sociology of Health & Illness*, *41*(S1), 98–115. https://doi.org/10.1111/1467-9566.12894

Popper, K. R. (2002). *Conjectures and refutations: The growth of scientific knowledge*. Routledge.

Prey, R. (2018). Nothing personal: Algorithmic individuation on music streaming platforms. *Media, Culture & Society*, *40*(7), 1086–1100. https://doi.org/10.1177/0163443717745147

Quine, W. V. (1951). Two Dogmas of Empiricism. *The Philosophical Review*, *60*(1), 20–43. https://doi.org/10.2307/2181906

Ribes, D., Jackson, S., Geiger, S., Burton, M., & Finholt, T. (2013). Artifacts that organize: Delegation in the distributed organization. *Information and Organization*, *23*(1), 1–14. https://doi.org/10.1016/j.infoandorg.2012.08.001

Ricœur, P. (1970). *Freud and philosophy: An essay on interpretation*. Yale Univ. Pr.

Ricœur, P. (2006). *The rule of metaphor: The creation of meaning in language* (Repr). Routledge.

Robbins, S. (2019). A Misdirected Principle with a Catch: Explicability for AI. *Minds and Machines*, *29*(4), 495–514. https://doi.org/10.1007/s11023-019-09509-3

Robinson, O. C. (2011). The Idiographic / Nomothetic Dichotomy: Tracing Historical Origins of Contemporary Confusions. *History & Philosophy of Psychology*, *13*(2), 32–39.

Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton University Press.

Rosch, E. (2011). "Slow Lettuce": Categories, Concepts, Fuzzy Sets, and Logical Deduction. In R. Belohlavek & G. Klir (Eds.), *Concepts and Fuzzy Logic* (pp. 89–120).

Rosner, R. I. (2014). The "Splendid Isolation" of Aaron T. Beck. *Isis*, *105*(4), 734–758. https://doi.org/10.1086/679421

Ruckenstein, M. (2014a). Visualized and Interacted Life: Personal Analytics and Engagements with Data Doubles. *Societies*, *4*(1), 68–84. https://doi.org/10.3390/soc4010068

Ruckenstein, M. (2014b). Visualized and interacted life: Personal analytics and engagements with data doubles. *Societies*, *4*(1), 68–84.

Rüppel, J. (2019). "Now Is a Time for Optimism": The Politics of Personalized Medicine in Mental Health Research. *Science, Technology, & Human Values*, *44*(4), 581–611. https://doi.org/10.1177/0162243919845049

Russell, B. (1920). *Introduction to Mathematical Philosophy* (2nd ed.). The Macmillon Co.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7–55). Elsevier.

Salvatore, S., Gelo, O., Gennaro, A., Manzo, S., & Al Radaideh, A. (2010). Looking at the Psychotherapy Process as an Intersubjective Dynamic of Meaning-Making: A Case Study with Discourse Flow Analysis. *Journal of Constructivist Psychology*, *23*(3), 195–230. https://doi.org/10.1080/10720531003765981

Salvatore, S., & Valsiner, J. (2008). Idiographic Science on its Way: Towards Making Sense of Psychology. *Yearbook of Idiographic Science*, *1*, 9–19.

Salvatore, S., & Valsiner, J. (2010). Between the General and the Unique: Overcoming the Nomothetic versus Idiographic Opposition. *Theory & Psychology*, *20*(6), 817–833. https://doi.org/10.1177/0959354310381156

Sanches, P., Howell, N., Tsaknaki, V., Jenkins, T., & Helms, K. (2022). Diffraction-in-action: Designerly Explorations of Agential Realism Through Lived Data. *CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3491102.3502029

Sanz, M. F., Acha, B. V., & García, M. F. (2021). Co-Design for People-Centred Care Digital Solutions: A Literature Review. *International Journal of Integrated Care*, *21*(2), 16. https://doi.org/10.5334/ijic.5573

Schafer, R. (1985). Wild Analysis. *Journal of the American Psychoanalytic Association*, *33*(2), 275–299. https://doi.org/10.1177/000306518503300201

Schiffrin, D. (1987). *Discourse Markers* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511611841

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Schüll, N. D. (2016). Data for life: Wearable technology and the design of self-care. *BioSocieties*, *11*(3), 317–333. https://doi.org/10.1057/biosoc.2015.47

Schwennesen, N. (2019). Algorithmic assemblages of care: Imaginaries, epistemologies and repair work. *Sociology of Health & Illness*, *41*(S1), 176–192. https://doi.org/10.1111/1467-9566.12900

Seaver, N. (2011). "This Is Not a Copy": Mechanical Fidelity and the Re-Enacting Piano. *Differences*, *22*(2–3), 54–73. https://doi.org/10.1215/10407391-1428843

Seaver, N. (2015). The nice thing about context is that everyone has it. *Media, Culture & Society*, *37*(7), 1101–1109. https://doi.org/10.1177/0163443715594102

Semel, B. M. (2021). Listening Like a Computer: Attentional Tensions and Mechanized Care in Psychiatric Digital Phenotyping. *Science, Technology, & Human Values*, 01622439211026371. https://doi.org/10.1177/01622439211026371

Sharon, T., & Zandbergen, D. (2017). From data fetishism to quantifying selves: Self-tracking practices and the other values of data. *New Media & Society*, *19*(11), 1695–1709. https://doi.org/10.1177/1461444816636090

Shedler, J. (2011). Science or ideology? *American Psychologist*, *66*(2), 152–154. https://doi.org/10.1037/a0022654

Silverman, D., Bor, R., Miller, R., & Goldman, E. (2004). 'Obviously the advice is then to keep to safer sex': Advice-giving and advice reception in AIDS counselling. In *AIDS: rights, risk and reason* (pp. 182–199). Taylor & Francis.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, *23*(3–4), 193–229. https://doi.org/10.1016/S0271-5309(03)00013-2

Simmons, K., & LeCouteur, A. (2011). 'Hypothetical active-voicing': Therapists 'modelling' of clients' future conversations in CBT interactions. *Journal of Pragmatics*, *43*(13), 3177–3192. https://doi.org/10.1016/j.pragma.2011.06.002

Sivakumar, A. (2023). Data Surrogates as Hosts: Politics of Environmental Governance. *Catalyst: Feminism, Theory, Technoscience*, *9*(1), Article 1. https://doi.org/10.28968/cftt.v9i1.38144

Skinner, B. F. (1965). *Science and human behavior* (First Free Press Paperback edition). The Free Press.

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: A systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology*, *43*(2), 476–493. https://doi.org/10.1093/ije/dyu038

Stern, D. B. (2002). Words and Wordlessness in the Psychoanalytic Situation. *Journal of the American Psychoanalytic Association*, *50*(1), 221–247. https://doi.org/10.1177/00030651020500011201

Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge university press.

Sullivan, M. (1986). In what sense is contemporary medicine dualistic? *Culture, Medicine and Psychiatry*, *10*(4), 331–350. https://doi.org/10.1007/BF00049269

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Thornton, T. (2007). Should comprehensive diagnosis include idiographic understanding? *Medicine, Health Care and Philosophy*, *11*(3), 293. https://doi.org/10.1007/s11019-007-9117-8

Thyloth, M., Singh, H., & Subramanian, V. (2016). Increasing Burden of Mental Illnesses Across the Globe: Current Status. *Indian Journal of Social Psychiatry*, *32*(3), 254. https://doi.org/10.4103/0971-9962.193208

Tomičić, A., Malešević, A., & Čartolovni, A. (2021). Ethical, Legal and Social Issues of Digital Phenotyping as a Future Solution for Present-Day Challenges: A Scoping Review. *Science and Engineering Ethics*, *28*(1), 1. https://doi.org/10.1007/s11948-021-00354-1

Torous, J., Staples, P., Barnett, I., Sandoval, L. R., Keshavan, M., & Onnela, J.-P. (2018). Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. *Npj Digital Medicine*, *1*(1), Article 1. https://doi.org/10.1038/s41746-018-0022-8

Verhagen, S., van Os, J., & Delespaul, P. (2022). 5—Ecological momentary assessment and other digital technologies for capturing daily life in mental health. In D. J. Stein, N. A. Fineberg, & S. R. Chamberlain (Eds.), *Mental Health in a Digital World* (pp. 81–108). Academic Press. https://doi.org/10.1016/B978-0-12-822201-0.00017-4

Vertesi, J. (2012). Seeing like a Rover: Visualization, embodiment, and interaction on the Mars Exploration Rover Mission. *Social Studies of Science*, *42*(3), 393–414. https://doi.org/10.1177/0306312712444645

Watson, A., & Lupton, D. (2022). Remote Fieldwork in Homes During the COVID-19 Pandemic: Video-Call Ethnography and Map Drawing Methods. *International Journal of Qualitative Methods*, *21*, 16094069221078376. https://doi.org/10.1177/16094069221078376

Watson, J. B. (1970). *Behaviorism*. Norton.

Wenger, E. (1990). *Toward a theory of cultural transparency: Elements of a social discourse ofthe visible and the invisible* [Doctoral Dissertation]. University of California, Irvine.

Wheeler, G., Mills, N., Ankeny, U., Howsley, P., Bartlett, C., Elphick, H., & Dimitri, P. (2022). Meaningful involvement of children and young people in health technology development. *Journal of Medical Engineering & Technology*, *46*(6), 462–471. https://doi.org/10.1080/03091902.2022.2089252

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., Johns, N., Burstein, R., Murray, C. J., & Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet*, *382*(9904), 1575–1586. https://doi.org/10.1016/S0140-6736(13)61611-6

Wilson, M. (1993). DSM-III and the transformation of American psychiatry: A history. *American Journal of Psychiatry*, *150*(3), 399–410. https://doi.org/10.1176/ajp.150.3.399

Windelband, W. (1998). History and Natural Science. *Theory & Psychology*, *8*(1), 5–22. https://doi.org/10.1177/0959354398081001

Winthereik, B. R., & Langstrup, H. (2010). When patients care (too much) for information. In A. Mol, I. Moser, & J. Pols (Eds.), *Care In Practice: On Tinkering in Clinics, Homes and Farms* (pp. 195–214). Transaction Publishers.

Wittgenstein, L. (1968). *Philosophical investigations*. Basil Blackwell.

Zuckerman, C. H. P. (2021). Figure Composition. *Signs and Society*, *9*(3), 263–299. https://doi.org/10.1086/715515