

A. Foundations

This dissertation opens with a part that establishes the necessary foundations to grasp and engage with the research presented in the following. To achieve this objective, it is subdivided into two sections which aim to motivate the research and that present its conceptual background.

First, Section A.I provides the motivation for this research, presents evident research gaps and derives relevant research questions to be addressed. Subsequently, it outlines the general structure of this thesis before analyzing its broader positioning and discussing anticipated contributions.

Next, Section A.II introduces the conceptual foundations of this dissertation. Initially, it presents Embodied Interaction as a stance on technological design in which meaning arises through engaged interaction with our environment and technological artifacts. It then provides insights into recent developments in cognitive science around 4E Cognition to shed light on the role of the human body in sense-making and social interaction. Finally, it relates these two perspectives to the application domain of the research in this thesis, i.e., psychotherapy, and introduces the publicly funded research project MITMed as the research context of two of the studies included in this thesis.

I. Introduction

The introductory sections serve to motivate the research conducted in this dissertation and outline its structure and content. They especially highlight the relevance of the engaged interaction with the technological artifacts that increasingly permeate our everyday lives. Additionally, they explore the role of our bodies in making sense of our own and our counterpart's experiences of affect in social interaction, especially when our interactions are mediated by digital technology.

To this end, Section I.1 establishes the motivation for studying technologically mediated social interactions from an interdisciplinary perspective. Next, Section I.2 highlights persistent gaps in our understanding of such interactions and introduces the overarching research questions derived to address these gaps in current scholarly work. Section I.3 moves on to describing this dissertation's structure, before Section I.4 discusses its overall research positioning and design. Finally, Section I.5 presents the anticipated contributions of the dissertation.

I.1 Motivation

This dissertation starts from the premise that information technology (IT) has become a ubiquitous aspect of our everyday lives, essentially permeating the entirety of human experience and becoming ever more intertwined with our personal or work-related social interaction routines (Baskerville et al., 2020; Carter & Grover, 2015; Overby et al., 2010; Yoo, 2010). We schedule meetings in and are reminded of important upcoming dates by digital calendars, retrieve information on current societally and politically relevant events in online newspapers, consume media such as books and films on e-readers and tablets, keep our friends and family updated via social media, and use our smartphones to get directions to our favorite restaurant.

Perhaps the most evident way in which digital technology reshapes our social interactions is by directly altering the way in which we communicate and engage with our peers. Technological progress has enabled us to interact with one another using digital means, such as text messaging and e-mail for asynchronous or video conferencing calls for synchronous interactions. Such digital innovation, of course, offers the advantage of bridging geographical distance, increasing speed and reach as well as reducing the costs of interaction across locations, thus contributing to the pervasiveness of technologically mediated social interactions. Early research has shown how different forms of digital media and their technological capabilities enable human-to-human communication that is suitable for various interaction tasks. For instance, it was shown that communication media are chosen depending on the degree of information richness they afford; consider how a phone call can convey more informational depth than an uncommented standardized textual report, and might therefore be chosen for a specific task that requires

such depth (Daft et al., 1987; Daft & Lengel, 1986). More recent research has expanded on these perspectives, adding that the fit of “information transmission and information processing needs” (Dennis et al., 2008, p. 581) will influence interaction synchronicity, i.e., whether or not two interacting individuals simultaneously collaborate with a common focus on some shared task, as well as the choice and use of specific communication media.

These investigations have contributed to a profound understanding of how information is exchanged and processed in technologically mediated social interactions. However, they offer limited insights into the relational processes at play in these interactions. Crucially, despite the advantages of technological mediation, its use also entails considerable challenges for the quality of our social interactions and for the relationships formed within these contexts. An essential process underlying the development of such relationships is affective attunement, i.e., the degree to which two interactors continuously recognize and (adequately) respond to each other’s mental states (see below). Importantly, this process is contingent on interactors’ nonverbal synchrony, that is “the mutual temporal coordination in changes occurring in patients and therapists during treatment” (Jennissen et al., 2025, p. 1214). Put simply, in conventional face-to-face interactions, interactors do not merely rely on the content of verbally exchanged information, but also on nonverbal, embodied¹ cues such as gestures and facial expressions. Such behavior is a central prerequisite for affective attunement because it shapes how the exchanged information is received and interpreted. Prior research has consistently shown the relevance of embodied communication for our abilities to understand each other’s mental states and coordinate behavior in social interactions (García et al., 2022; Hall et al., 2019; Krueger, 2018; Vogeley, 2018). In turn, affective attunement is a precondition for the emergence of a shared social understanding of each other and the situation, and thus for joint sense-making processes to occur during the interaction (De Jaegher, 2018; De Jaegher & Di Paolo, 2007; García et al., 2022).

In technologically mediated social interactions, the transmission of these cues is often heavily impeded. While technological progress allows for interactions that become ever more similar to the “gold standard” of face-to-face interactions (e.g., video calls that allow you to see your interaction counterpart instead of just hearing their voice during a phone call), it remains difficult to argue that they are perfect substitutes (Stieger et al., 2023). Note for instance how, in video conferencing, the view of one’s counterpart is usually constrained to a fixed frame of the other’s face (or upper body) and does not allow for a full assessment of their body language (e.g., gestures, posture). Technological

¹ For now, assume the (naïve) definition that *being embodied* means that a phenomenon has a physical manifestation in the real world (Dourish, 2001); e.g., nonverbal communication as being manifest in bodily communication cues such as gestures or facial expressions. Section II.1 will elaborate in more detail about notions of embodiment.

challenges, such as low internet bandwidth, poor video and audio quality, or delays in transmission further complicate the assessment of the other's facial expressions or vocal cues (Almathami et al., 2020; Gullslett et al., 2021; Tulu & Chatterjee, 2008). Crucially, the severity of this interactional loss does not stem from the communication media not meeting demands for *informational* but for *affective* interaction processes. The attenuation of nonverbal communication cues in technologically mediated social interaction contexts therefore may pose a threat to the quality and depth of our interactions.

The problems associated with this interactional loss quickly became visible at a global scale during the COVID-19 pandemic. As a result of governments' social distancing measures to contain the spread of the virus, social interactions shifted toward technologically mediated contexts such as social media, messaging applications and video conferencing to an unprecedented degree (Drouin et al., 2020; Nguyen et al., 2020, 2022). Prior research has shown how the technological mediation of everyday social interactions has had negative effects on individuals' mental health and emotional well-being relative to face-to-face interactions (Stieger et al., 2023). Importantly, however, what appears problematic in individuals' private social interactions becomes even more consequential in other social interaction contexts, especially those where social understanding and affective attunement are central determinants of interaction outcomes. These contexts are characterized not merely by a need for information exchange, but also by a need for continuous relationship work between the interactors. The healthcare sector is representative of such contexts since its outcomes are particularly contingent on the quality of the relationship established within the clinical interaction between a healthcare professional and their patient (Kelley et al., 2014), which in turn is largely built on affective attunement and mutual trust (Paul & McDaniel, 2004). The research presented in this dissertation investigates the case of psychotherapy, a specific subdomain of the healthcare sector dealing with the treatment of mental disorders, as an illustrative context in which the concepts discussed above (i.e., nonverbal, embodied communication; affective attunement; relationship quality; interaction or treatment outcomes) become interlinked.

It is worthwhile, at this point, to briefly describe the relevance of treating mental disorders, both for individuals' well-being and for nation states and the (global) economy. According to recent estimates provided by the World Health Organization (WHO), over one billion people worldwide are affected by mental disorders, with depressive and anxiety disorders being the most prevalent (WHO, 2025). Consequences on the individual level include the inability to work, subjective suffering and perceptions of social exclusion (Kegel et al., 2022). Moreover, mental disorders represent an enormous stress factor for economies worldwide, with the annual global productivity losses resulting from treating schizophrenia, depression and anxiety disorders alone amounting to roughly one trillion

USD (WHO, 2025). Yet, those affected are often heavily underserved with healthcare services: fewer than 10% of individuals suffering from depressive disorders currently receive formal treatment (WHO, 2025), highlighting a substantial gap in coverage.

A promising concept to address this gap is telemedicine, which generally refers to the delivery of healthcare services over a geographic distance using technology (Bashshur, 1995). Not restricted to mental healthcare, telemedicine can increase access to healthcare services previously unavailable to individuals and contributing to the reduction of healthcare access inequalities (Hollander & Carr, 2020; Hwang et al., 2022). Notably, the COVID-19 pandemic substantially increased existing pressures on the adequate delivery of healthcare services since it (and the related policy countermeasures) led to additional spikes in prevalence and symptom severity of mental disorders (Cao et al., 2020; Peters et al., 2020). Thus, together with the policy measures implemented by national governments, COVID-19 drove psychotherapists worldwide to switch to technology-mediated interactions such as video consultations in order to continuously provide treatment to patients suffering from mental disorders. Prior research shows how delivering technologically mediated psychotherapy became the *de facto* standard during the pandemic as therapists were practically forced to provide remote therapy through information and communication technology so as to guarantee continuity of care (Markowitz et al., 2021; Simpson et al., 2021). This switch to technologically mediated clinical interactions substantially accelerated the adoption of digital technologies for healthcare documented previously (Botrugno, 2021), and that of telemedicine services in particular, which had traditionally been adopted only slowly (Bashshur et al., 2020).

Importantly, the widespread adoption of video consultations as a mode of psychotherapy delivery essentially preserves the fundamentally interactional characteristic of psychotherapeutic interventions: whether it is delivered face-to-face or using technological mediation, psychotherapy remains a treatment built mostly around communication, where the social interaction between therapist and patient forms the foundation for therapeutic change (Jennissen et al., 2025). In this interaction setting, the quality of the relationship between both interactors maintains a crucial role. Research has consistently shown it to be a significant determinant of therapy success (Asay & Lambert, 1999; Flückiger et al., 2018; Horvath et al., 2011). As will be discussed later on, the establishment and maintenance of this relationship critically hinges on the therapist's ability to detect, acknowledge, and repair ruptures (or breakdowns) of the relationship (Safran et al., 2011). This (rupture-repair) process, in turn, is linked to individual and joint or *participatory sense-making*, i.e., the emergence of a shared social understanding (e.g., about mental states) arising from repeated interactions and affective attunement based in part on nonverbal, bodily communication (De Jaegher & Di Paolo, 2007; Deres-Cohen et al., 2021; Foley & Gentile, 2010; García et al., 2022). Figure 1 contains a simplified depiction of the relationship between the interactors within the psychotherapy dyad, the

role of nonverbal, bodily communication cues, affective attunement through rupture-repair processes, the patient-doctor relationship, and psychotherapy outcomes. Note that the relations in this figure initially hold regardless of whether or not the clinical interaction is technologically mediated.

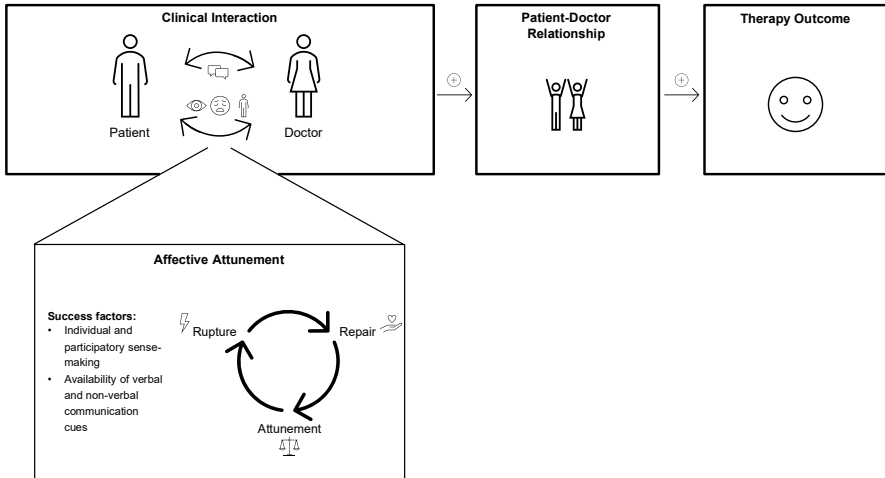


Figure 1. The Role of Affective Attunement in Psychotherapy

As established, when the clinical interaction is technologically mediated, the availability of crucial nonverbal, embodied communication cues is heavily reduced (Cataldo et al., 2021; Roesler, 2017). As a result, therapists continue to note persisting concerns regarding their ability to develop strong relationships with their patients in such interaction contexts (Simpson et al., 2021), which in turn may lead to deteriorations of the patient-doctor relationship and, thereby, threaten therapy success.

Understanding how digital technology might be designed and used to allow for more affective attunement and participatory sense-making in technologically mediated interactions currently remains an open question at the intersection of information systems (IS) and human-computer interaction (HCI), and is particularly relevant in clinical practice. This dissertation aims to address that gap by investigating different strategies for how technologically mediated interactions can be better suited to the needs of clinical interactions, particularly in psychotherapy: first, it explores the design of virtual reality (VR) environments for clinical interactions, thus replacing the conventional delivery mode of video consultations; second, it draws on extensive research efforts from a three-and-a-half year publicly funded research project (MITMed, see Section II.3.4) aimed at technologically enhancing, rather than replacing, video consultations by introducing an innovative nonverbal communication channel to transmit vibrational cues in real-time. Overall, this dissertation draws on conceptual foundations from the HCI literature and

cognitive sciences (see Sections II.1 and II.2, respectively). It thereby argues for the importance of participatory sense-making in technologically mediated clinical interactions and presents empirically grounded design knowledge to support this important aspect of social interactions.

I.2 Research Gaps and Research Questions

The key objective of this dissertation is to contribute to an understanding of which and how technologies can be used to support technologically mediated clinical interactions, especially in the context of psychotherapy. Focusing on potential contributions of technology is imperative given the pervasiveness of digital technology in every aspect of our lives as well as the ever-increasing digitalization of the healthcare domain (Botrugno, 2021). While, from an idealist perspective, one might want to argue to perform every interaction in a “gold-standard” face-to-face context, this is simply not feasible given the resource constraints and accessibility issues we regularly face in the healthcare domain, e.g., regarding specialist availability and rural patients’ long travel times (Hwang et al., 2022).

A central argument that this dissertation builds on is that research to date has largely conceptualized technology employed in social interaction based on the degree of richness in the information conveyed (Daft et al., 1987; Dennis et al., 2008), thus reducing the investigation down to the level of technological properties (Schultze & Brooks, 2019). As a result, research into the technology-enabled remote delivery of healthcare services has taken on a similarly information-focused perspective, in which interactions depend on the exchange and processing of information by both interactors (Paul, 2006; Serrano & Karahanna, 2016). This view reflects a relative neglect of the role of embodied communication elements in current technological designs that might provide more opportunities for meaningful social interactions. Additionally, this perspective implicitly builds on simplistic assumptions of technologically mediated interaction as being by definition “disembodied” (Vidolov, 2024), thereby overlooking (1) the capabilities provided, for instance, by immersive virtual environments as sites of social interaction, and (2) the role of the human body and nonverbal communication in social interactions more generally. This dissertation turns to two research strands to problematize these assumptions. First, it will draw on insights from the HCI literature on individuals’ engaged interaction with a technologically permeated environment as a source of meaning (see Section II.1). Second, it will consider insights from cognitive science regarding the relevance of the human body in social interactions and affective attunement (see Section II.2).

As noted, major concerns regarding therapists’ capabilities of building strong relationships with their patients persist because of the reduced availability of nonverbal,

bodily communication cues provided in technologically mediated interactions. These concerns reflect therapists' fear that the informational richness conveyed in these interactions, e.g., in video consultations, is insufficient for the case of psychotherapy (García et al., 2022; Simpson et al., 2021). Thus, on the most abstract level, the research in this thesis is driven by the question of **how digital technology designs can be used to support affective attunement and participatory sense-making in technologically mediated clinical interactions**, focusing on the application domain of digital psychotherapy. To address this overarching research question (RQ), this thesis explores two possible strategies for technological designs to facilitate participatory sense-making whenever face-to-face interactions are not feasible. These strategies are described in the following.

A first strategy could be to move away from the idea of video consultations for healthcare, and psychotherapy, and consider **entirely different forms of technological mediation for meaningful clinical interactions**. As will be discussed in more detail later on, despite their proven clinical effectiveness (Almathami et al., 2020), video consultations do not seem to provide the same opportunities for building strong relationships between interaction partners as do face-to-face interactions (Norwood et al., 2018; Simpson & Reid, 2014). Therefore, among multiple other reasons, the uptake of video consultations has traditionally been slow in psychotherapy (Simpson et al., 2021), at least until the pandemic practically forced therapists to employ them to maintain continuity of care. Still, other technological solutions could be employed to provide for more meaningful clinical interactions in digital psychotherapy instead. One objective of this dissertation is therefore concerned with the exploration of such alternative technological designs (i.e., beyond video consultations) for digital psychotherapy.

A prime candidate technology that might be considered a suitable alternative to video consultations in this context is virtual reality. Essentially, VR enables users to be present in fully digital, interactive worlds that afford opportunities to interact with others and the (digital) environment (Dourish, 2001; Wohlgenannt et al., 2020). A key to the successful implementation and use of VR interaction environments, and simultaneously their main advantage compared to video consultations, is their capacity to establish perceptions of presence for their users (Schultze & Brooks, 2019; Schultze & Orlikowski, 2010). Feeling present in VR can mean a lot of different things, and definitions are not always well-delineated; however, in the context of this thesis, presence is understood from an interactional perspective. This view describes presence as the degree to which individuals feel involved in a temporally and spatially bounded setting in which they cooperate cognitively and emotionally with others (that are equally present) in a joint activity (Schultze & Brooks, 2019). Importantly, such presence does not emerge based on users finding themselves in VR environments per se. Rather, it is enacted through interaction forms that integrate nonverbal, bodily communication cues such as common

spatial orientation, gestures or facial expressions, that can be employed in VR environments more easily than in video consultations (Schultze & Orlikowski, 2010; Wohlgenannt et al., 2020). These technological opportunities and their effects on the potential emergence of perceptions of social presence might thereby facilitate processes of affective attunement and participatory sense-making that are particularly relevant in healthcare and, specifically, in (digital) psychotherapy (see Section II.3.2). Applications of VR thus seem particularly promising for digital psychotherapy. However, current knowledge on the design of such applications remains limited, and most VR applications are disorder-specific (Bermúdez i Badia et al., 2019; Valmaggia et al., 2016). While Turan Akdag et al. (2023) designed transdiagnostic VR therapy environments, these were constrained by technological limitations of using pre-existing VR applications and their capacities of establishing a sense of presence. Taken together, this highlights a gap in (IS) research regarding the design of flexible VR environments for digital psychotherapy that leverage the technology's benefits across patient conditions. Therefore, this dissertation addresses the following research question:

RQ 1: How can virtual reality environments be designed to support patients' and therapists' interactional needs in technologically mediated psychotherapy?

Despite the promise that VR environments entail for the technological mediation of clinical interactions, it would be challenging to argue for the complete replacement of traditional video consultations for the delivery of mental healthcare through VR. On the one hand, implementation costs would be high (e.g., regarding novel VR hardware). On the other, there is ample evidence for the general effectiveness of video consultations and their potential contributions towards closing persistent gaps in mental health service coverage, especially with the experiences made during the pandemic (Almathami et al., 2020; Gullslett et al., 2021). Still, prior research has repeatedly highlighted interactors' concerns regarding this form of telemedicine, especially on the therapist side. Therapists seem particularly concerned about the shift of psychotherapeutic treatment into a technologically mediated environment such as video consultations due to the medium's perceived shortcoming in providing an environment suitable for the establishment of social presence and trusting relationships (Cataldo et al., 2021; Simpson et al., 2021). This finding is in line with other research indicating health care professionals' resistance toward novel technology, more generally (Kumar et al., 2022; Walter & Succi Lopez, 2008; Y. Yang et al., 2024). Against this background, and given that therapists might have recently become more accustomed to using video consultations to deliver mental healthcare, exploring the ways in which traditional video consultation systems might be technologically enhanced, rather than replaced, might represent another sensible way of addressing the overall problem. Therefore, the second strategy explored in this dissertation encompasses investigations of **how such technological enhancement**

might be realized to support affective attunement and participatory sense-making in video consultations.

To this end, this dissertation again turns to the relevance of embodied communication cues to foster affective attunement. Currently, there seems to exist no research regarding potential technological augmentations to the video consultation context by increasing the availability of nonverbal cues (regardless of whether one assesses the IS, HCI, or psychotherapy literature). This is surprising since the technological shortcomings associated with traditional video consultations—such as fixed camera frames, poor video and audio quality resulting from subpar equipment or low-bandwidth internet connections—are well documented in prior research (Almathami et al., 2020; Gullslett et al., 2021; Markowitz et al., 2021; Tulu & Chatterjee, 2008). Most often, research has provided rather general advice on how interactors might adapt to the video consultation setting. This includes, for instance, recommendations on camera positioning, lighting, general therapy-room setup and individuals' distance to their camera and computer screens, or on making adjustments to explicitly verbal communication behaviors so as to compensate for the lack of nonverbal, bodily communication cues (Markowitz et al., 2021; Shore, 2013; Simpson et al., 2021). Overall, however, one has to note the surprising shortcoming of IS (design) research acknowledging and accounting for the role that the human body might play as an active participant in technologically mediated social interactions based on affective attunement and participatory sense-making. While there seems to exist an agreement on the relevance of nonverbal, bodily communication for successful affective attunement in clinical interactions, we lack an understanding of how such forms of embodied interaction with systems and with other participants of the interaction might be reintroduced within the constraints of video consultations. Crucially, this dissertation assumes that addressing this gap requires investigations of two separate but equally relevant aspects.

First, it remains unclear how the introduction of an enhanced video consultation mode would affect prospective users' (i.e., therapists' and patients') attitudes and usage intentions relative to traditional video consultations. While enhancements of video consultations (e.g., through embodied communication cues) might positively influence affective attunement, prior research has repeatedly documented clinicians' general skepticism regarding the introduction of novel technology (Kumar et al., 2022; Walter & Succi Lopez, 2008). Thus, before targeting the generation of more explicit design knowledge regarding the technological enhancement, an investigation of users' perceptions of an enhanced video consultation mode seems imperative. Therapists' perceptions are especially relevant in this regard, since they would assume a gatekeeper role should such a novel system be implemented into clinical practice. After all, it is initially the therapists' role to establish a given mode of technological mediation as a suitable place in which to practice therapeutic work (e.g., through the practices noted by Simpson