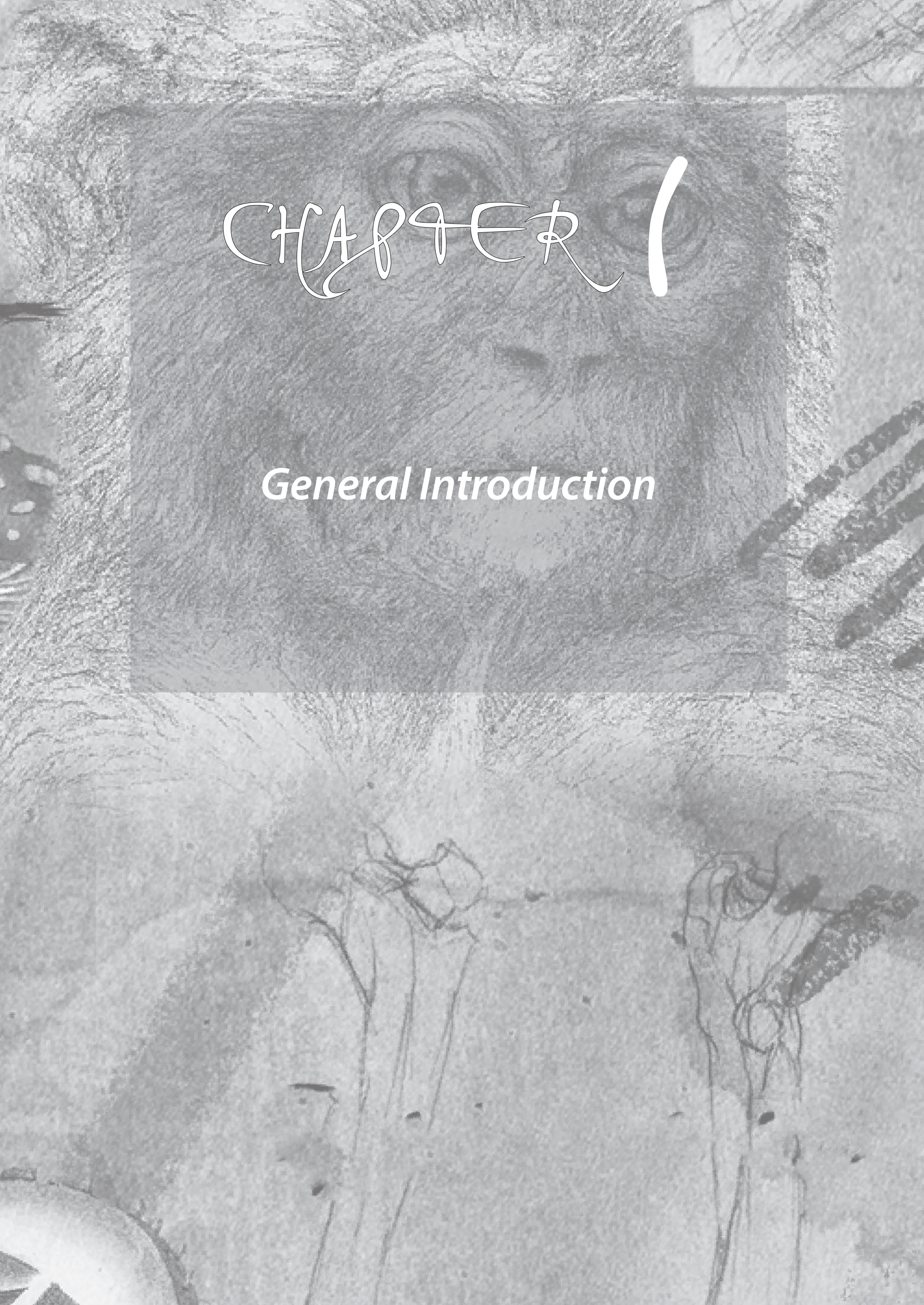
Interactional beings: the power of automatic mimicry and nonverbal cues in shaping human-human and human-robot naturalistic interactions
Diana, F.



CHAPTER 1

General Introduction

Life, even in its earliest forms, could not exist without interaction. As the biologist Rachel Carson beautifully remarked: "*In nature, nothing exists alone*" (Carson, 2000). This is especially true for species living in communities, where the immediate environment is surrounded by conspecifics with whom interaction is inevitable. Ants share pheromonal trails to coordinate their foraging efforts (Czaczkes et al., 2012); pilot whales synchronize their movements when avoiding predators (Senigaglia et al., 2012); primates groom each other not just for hygiene, but to build and maintain social bonds (Dunbar, 1991). Across species, social interaction is a fundamental requirement for survival.

Humans, too, are deeply social. Like other species with a complex social structure, we rely on various strategies to make sense of others and navigate our social environment. Among the most important is the exchange of nonverbal emotional expressions and cues, which can convey rich information about others' internal state (Patterson, 2017). Emotional cues may be produced with communicative intent or unconsciously triggered in response to emotion-eliciting events, for instance, through behavioral and physiological changes (Keltner et al., 2019; Kret, 2015). Particularly relevant for a smooth interaction is the automatic, often unconscious mimicry of these cues, a phenomenon linked to interpersonal trust, affiliation, and empathy (Hale, 2017; Lakin, 2013; Preston & Waal, 2002). Such mimicry can also occur at the autonomic level – for example, in the synchronization of heart rate or skin conductance – which is likewise associated with emotional understanding and prosociality (Behrens et al., 2020; Järvelä et al., 2014). Yet, is mimicry *always* affiliative or prosocial? After all, not all nonverbal cues we produce are meant to connect; people scratch, yawn, or fidget not to build rapport but for entirely different reasons, and we may smile not only to express sincere happiness, but also to hide our true feelings. Will the effects of nonverbal cues and their mimicry *still* be prosocial when the context, the intention, or the meaning of the cue itself varies?

This question becomes even more tangled as, unlike other species, the landscape of human interaction keeps expanding. Our social environment is increasingly shaped by technology: from virtual communication via video call to interactions with artificial agents (i.e., avatars, chatbots, physical robots), these technologies now occupy spaces once exclusively dedicated to humans or other animals. Do the same cues and mechanisms that evolved for flesh-and-blood human interaction still function in these novel digital contexts? Further complicating the matter, digital technologies – as social cues in general – are not interpreted uniformly across cultures (Kawahara et al., 2021; Lim et al., 2021), questioning whether cues, especially when equipped to artificial agents, are perceived and acted upon differently among cultures.

The present dissertation investigates if and how social cues and their mimicry shape prosocial behavior, and how the effects depend on the context where the interaction occurs and the specific cue involved. In a series of naturalistic real-life studies in heterogeneous samples, I trace a line from direct human-human interaction to video-mediated communication, to encounters with artificial agents such as avatars and robots, whilst considering the cultural lenses through which technologies and cues are perceived. Across six chapters, I argue that neither the effects of social cues nor of their mimicry can be understood in isolation from the conditions in which they emerge; rather, they are adaptive tools, shaped by the social context, the modality of interaction, the cultural background, and the type of agents involved. This chapter provides a comprehensive summary of the relevant literature and underlying rationale, creating an integrated overview that establishes the context and foundation for the chapters that follow in the dissertation.

The ubiquity and significance of nonverbal cues

Nonverbal cues are intrinsic to social interaction. From facial expressions and posture to tactile and chemical signals, these cues are widespread and highly adaptive across species, enabling animals to coordinate activities, avoid predators, form social bonds, and compete for resources. Nonverbal cues are generally a fast and efficient way to adapt to social and environmental changes by helping to anticipate others' actions and communicating information (De Waal & Preston, 2017). In humans, the variety of nonverbal cues exchanged during social interactions is remarkable, ranging from overt cues – such as facial expressions, body postures, gestures, and vocalization (Patterson, 2017) – to more subtle, covert ones, like changes in pupil size, facial reddening (Kret, 2015) and even scent (de Groot & Smeets, 2017).

While some cues can be consciously modulated, others are produced involuntarily (Kret, 2015) (Kret, 2015) and can even be perceived automatically, outside of conscious awareness (Tamietto et al., 2009). As a result, the back-and-forth of nonverbal behaviors requires little cognitive energy, even when verbal communication becomes more demanding. This automaticity makes nonverbal cues highly efficient and persistently active: as long as some form of sensory input – be it visual, auditory, or olfactory – is available, the nonverbal channel is unlocked (Patterson, 2017). For these qualities, the continuous flow of cues that accompanies social interactions serves multiple adaptive functions. As a start, it conveys critical information about the emotional state and intentions of others: a smile can signal happiness (Ekman et al., 2017) or the desire to affiliate (Crivelli & Fridlund, 2019), or changes in pupil size can indicate alertness, interest, surprise, cognitive effort as well as other emotional states (Kret, 2018b; Murphy et al., 2014; Preuschoff et al., 2011). The information embedded in cues like facial expression can help the observer to anticipate others' future behavior and changes in the surrounding environment (Morimoto & Fujita, 2012; Waller et al., 2017), with crucial advantages for survival. Nonverbal cues also shape interpersonal judgment and guide our intuitions. Subtle cues like gaze direction, micro-expressions, or pupil size can signal trustworthiness or deceit, helping us to rapidly assess potential allies or threats (Kret & De Dreu, 2017; Todorov, 2008). Leveraging basic evolutionary threat-detection mechanisms (Delgado et al., 2006), we can automatically evaluate the trustworthiness of a stranger within milliseconds based on their facial configuration (Willis & Todorov, 2006) or pupil size (Frisanco et al., 2023; Kret et al., 2015; Wehebrink et al., 2018a). Importantly, these cues not only allow us to detect trustworthy partners, they also guide our own prosocial behavior (**Box 1**) (van Breen et al., 2018), a core topic of the present dissertation. A crucial feature of nonverbal cues is that their production and perception can occur simultaneously. As in, individuals unconsciously mimic each other's behavior. This alignment permeates social interactions and is critical to coordinate actions with those around us.

Automatic mimicry

It is not easy to imagine a social interaction where we do not mirror each other's behaviors. This mirroring, better known as automatic mimicry, occurs instinctively, often without intention or awareness, and it is as important to social animals as nonverbal cues themselves. Scratching when observing someone else doing it, laughing when others do, or yawning after someone else nearby yawns are just a few examples of automatic mimicry in everyday interactions. While the traditional definition of automatic mimicry focuses only on the unconscious imitation of overt behavior like facial expressions or postures (Chartrand & van Baaren, 2009), humans also mimic on a deeper bodily level. This physiological synchrony, the temporal alignment of autonomic activity, can occur across multiple physiological

Box 1. Prosocial Behavior: Definitions and Dimensions

Prosocial behavior broadly refers to voluntary actions intended to benefit others rather than oneself (Batson & Powell, 2003). It includes a wide array of behaviors such as trust, honesty, cooperation, helping, altruism (costly helping) and sharing. In both humans and non-human animals, these behaviors play a key role in promoting social cohesion and group survival (Burkart et al., 2014; Tomasello & Vaish, 2013). Prosociality exists on a behavioral continuum that includes its functional opposite: dishonesty, free-riding, and other self-serving behaviors. Like prosocial behavior, anti-social behaviors have ancient evolutionary origins and adaptive purposes. From an evolutionary perspective, manipulating others through deception – rather than physical force – likely offered advantages in competing for resources, status, or mates, much like other deceptive strategies in the animal kingdom (e.g., camouflage or batesian mimicry) (Vasconcellos Lemos et al., 2019). In humans, decisions to behave prosocially (or not) are shaped by a complex interplay of cognitive processes – such as moral reasoning, cognitive cost, and cost-benefit analyses (Jacobsen et al., 2018) – and intuitive unconscious processes. These unconscious evaluations often depends on the exchange of nonverbal cues that can help us understanding others' intentions (Kret & De Dreu, 2019). Thus, nonverbal cues are highly relevant for understanding how prosocial behavior is shaped in real-life interaction. In the studies presented here, prosocial behavior is operationalized through economic games or behavioral tasks that capture trust, reciprocity, altruism, or (dis)honesty by simulating real-world decision making under uncertainty under varying cue and context conditions. These paradigms allow researchers to measure how individuals balance self-interest with the potential benefits of mutual cooperation.

signals and in diverse social contexts (Palumbo et al., 2017). There is an ongoing debate surrounding whether other forms of alignment (e.g., synchrony, true imitation, coordination, etc.) could fall under the umbrella of automatic mimicry. In the present dissertation, I take the definition formulated by Prochazkova and Kret (2017). Here, the authors use 'automatic mimicry' as a superordinate term comprising various types of synchronous behaviors that occur unconsciously at different levels (Prochazkova & Kret, 2017a). The authors make a distinction between mimicry involving muscle movements (motor mimicry) and mimicry of physiological patterns (autonomic mimicry). Motor mimicry results from two or more people mimicking each other's motor movements, such as facial expressions (Likowski et al., 2012), conversational gestures (Kimbara, 2006), body posture (Tia et al., 2011), laughter (Davila-Ross & Palagi, 2022), yawns (Massen & Gallup, 2017), and scratches (Schut et al., 2015). Autonomic mimicry involves any synchronous physiological fluctuation – at the level of heart rate (Riddoch & Cross, 2012), skin conductance (Behrens et al., 2020), pupil diameter (Kret et al., 2014), hormonal level (Liu et al., 2005), or brain activity (Schilbach & Redcay, 2025) – within the group or dyad.

Despite the growing literature, we still know relatively little about how these different levels of mimicry interact. It remains unclear whether autonomic responses precede and facilitate motor mimicry, or whether the reverse is true, or if both occur simultaneously or independently. Furthermore, synchrony and mimicry can be thought of as different ways of computing or conceptualizing similar phenomena. Synchrony is often assessed using continuous time-series data – for example, measuring the temporal alignment of heart rate or pupil diameter over time – whereas mimicry is typically captured through discrete point events, such as counting how often one participant scratches shortly after the other does. For this reason, in the present dissertation, I use the terms "mimicry" and "synchrony" interchangeably, depending on the computational or methodological approach used in each context.

Why do we mimic

Just like nonverbal cues, automatic mimicry has been observed in many species, including great apes (Kret & van Berlo, 2021), Old World monkeys (Whitehouse et al., 2016), meerkats (Palagi et al., 2019), lemurs (Valente et al., 2022), dogs (Palagi et al., 2015), pigs (Norscia et al., 2021), sun bears (Taylor et al., 2019), dolphins (Jaakkola et al., 2010), and budgies (Miller et al., 2012). Due to mimicry's prevalence and ancient phylogenetic roots, scientists from various fields have become increasingly interested in understanding its adaptive significance and evolutionary trajectory. In other words, *why* do we mimic? There have been several theories offering diverse perspectives on mimicry's function and evolutionary origins. A popular view on the function of automatic mimicry is the Social Glue hypothesis (Lakin et al., 2003), which lays its foundations on a distinction between the *original* function of mimicry and its function nowadays. The authors argue that mimicry may have initially originated as a mechanism to quickly gather information from the immediate environment: if I imitate your fear or vigilance responses, I will be quicker in detecting danger and more likely to survive. However, its core function has shifted, and mimicry acts now as a "social glue" to enhance affiliation, therewith increasing group cohesion (Lakin et al., 2003). This theory stems from the idea that humans, unlike many other species, have relatively fewer immediate physical threats (e.g., predators), and the original survival-related function of mimicry may be less relevant for humans today. Instead, mimicry in humans has evolved more for social purposes – to promote affiliation, cooperation, and group cohesion. The Social Glue hypothesis finds support in a substantial amount of studies showing that mimicry increases the likeability and trustworthiness of the partner (Lakin & Chartrand, 2003; Prochazkova et al., 2018a; Wessler et al., 2024), smoothens interactions (Fasya et al., 2024), and is more pronounced when there are pre-existing affiliative bonds, such as with ingroup members (van der Schalk et al., 2011). Research on non-human animals found complementary results, where mimicry seems to play a crucial role in prolonging and regulating a play session (Mancini et al., 2013) and to occur more often between socially close individuals (Norscia et al., 2021) or in more socially tolerant species (Scopa & Palagi, 2016). Beyond the social glue hypothesis, several other theories – across human and comparative research – have been formulated to explain why we mimic. For example, some authors proposed that interpersonal coordination, which includes different forms of mimicry, could be fundamental for pair bonding across species (Roth et al., 2021). Other authors think of mimicry as a social regulator, a social act dependent on an affiliative goal and on the emotional appraisal of the cue that is mimicked (Hess & Fischer, 2022). Mimicry has also been framed as a flexible social behavior that can be strategically deployed – without awareness – to maintain status, avoid conflict, or signal ingroup membership (Wang & Hamilton, 2012). While distinct, these theories revolve around a recurring theme: the role of mimicry in fostering affiliation, social bonds, and generally positive social consequences. Despite the convincing evidence supporting an affiliative consequences of mimicry, the empirical findings are far from uniform: for all the studies highlighting positive social outcomes of mimicry, there are just as many that point to contextual influence (Seibt et al., 2015), null (Hale & Hamilton, n.d.) or even negative effects of mimicry (van der Velde et al., 2010). Mimicry can undoubtedly foster affiliative interactions, but not always.

A disadvantage of the affiliative view of mimicry is that it overlooks the main characters of the act: the cues. Cues are not inherently affiliative. Their valence can span the entire spectrum from affiliative to antagonistic. While this has received less attention, research shows that we do, in fact, mimic negative expressions such as fear, disgust, and anger (Dijk et al., 2018; Rymarczyk et al., 2019) or, for non-human animals, anti-predator vigilance behaviors (Iki & Kutsukake, 2021). We also mimic cues with ambiguous valence. Behaviors like scratching and yawning illustrate this well: they are highly contagious, automatic, and context-dependent behaviors, whose mimicry occurs across a variety of scenarios, not limited to affiliative contexts. For instance, mimicry of scratching was observed among weakly bonded individuals, indicating tense situations (Laméris et al., 2020), but also in positive valence contexts like play (Neal & Caine, 2016). Contagious yawning can occur among individuals with strong social bonds (Norscia & Palagi, 2011), but also not (Valdivieso-Cortadella et al., 2023), and different yawning displays can signal both intimidation and benign intentions (Leone et al., 2014). Not only do we mimic cues with different valence, but we also mimic the same type of cue in different social contexts. Autonomic mimicry can emerge among teammates in cooperative situations (Romero-Martínez et al., 2019), as well as among opponents during competitions (Spapé et al., 2013). Further, mimicry of play faces during aggressive playfight can be accompanied by more offensive patterns, suggesting that even cues traditionally associated with positive affect can lead to differential outcomes (Cordoni et al., 2024).

The effects of mimicry seem to depend on what is mimicked and in which context. Since social cues often communicate internal states and intentions, mimicry has been proposed as a mechanism that facilitates the transmission of emotional states between individuals (Hatfield et al., 1993). Through the automatic replication of facial expressions, postures, or autonomic states, individuals may attune to the emotional states of others via sensorimotor and autonomic feedback (Prochazkova & Kret, 2017b). The attunement may occur independently of the emotional valence – whether it is joy and enthusiasm, but also fear, anger, or sadness – highlighting mimicry's role in emotional alignment, even though that alignment may not always be desirable. In this view, rather than being intrinsically affiliative, mimicry may function as a mechanism that helps individuals adjust to the dynamics of the conspecifics and the environment. Just as the context of interaction and its valence shape how a cue is perceived and responded to, the effects of mimicry may likewise depend on these factors. Genuine smiles, for instance, are associated with positive evaluations and their mimicry causally increases attraction (Arias-Sarah et al., 2024), but not always (Prochazkova et al., 2022), whereas mimicking angry faces, perceived as threatening, increases aversion (van der Velde et al., 2010). Zooming in more specifically on prosocial behavior (**Box 1**), mimicry has been shown to affect trust, honesty, and other forms of prosocial behavior, but its effects are strongly modulated by the characteristics of the cues being mimicked. Pupil size is an excellent example. Human faces with large pupil sizes are generally perceived as positively attractive and intelligent (Hess, 1975; Lau et al., 2022; Tombs & Silverman, 2004). In their presence, people tend to behave more honestly (van Breen et al., 2018) and show greater approach behavior (Brambilla et al., 2018). Consequently, the mimicry of large pupil size is associated with higher levels of trust (Kret & De Dreu, 2017; Prochazkova et al., 2018a). In contrast, small pupils can signal boredom or uninterest, are perceived as less attractive and intelligent (Lau et al., 2022; Tombs & Silverman, 2004), and trigger avoidance responses (Brambilla et al., 2019). In this case, mimicry can reduce trust (Wehebrink et al., 2018b). Importantly, the majority of the abovementioned studies are correlational, which limits conclusions about causality.

In sum, the last few paragraphs aimed to highlight the complexity of mimicry's functions and consequences, a topic I will touch upon throughout the dissertation. Specifically, I will argue that rather than serving a single affiliative purpose, mimicry emerges as a flexible mechanism that can help individuals align with one another emotionally and behaviorally, reinforcing bonds, regulating tension, or responding to threat. This alignment can have meaningful outcomes that either promote or inhibit prosocial behavior. As social interaction increasingly occurs in technologically mediated environments – whether through screens, avatars, or artificial agents – these core processes might be reshaped. The cues we perceive, the contexts in which mimicry occurs, and the outcomes it produces may differ substantially when humans interact through or with technology.

Mediated Interaction and the Technological Layer

The last few decades have witnessed unparalleled technological progress that has fundamentally reshaped our social fabric. From work meetings and mental health services to video-calling with your loved ones, many aspects of our social life have increasingly transitioned to digitally supported environments. While these technologies can be useful alternatives and, in some cases, effective, surprisingly little is known about what happens to the dynamic of social interactions through online video call platforms. Compared to other communication media (e.g., phone calls), video calls have the advantage of providing access to at least some nonverbal cues. That said, these cues are often not as straightforward as during a face-to-face interaction (Fauville et al., 2021). Factors such as camera placement, poor lighting, delayed transmission, or low-quality video can conceal or distort subtle facial expressions or cues like eye gaze and pupil dilation, making nonverbal cues harder to interpret during video interactions (Duffy & Benotsch, 2025). This often makes individuals more self-conscious about their nonverbal behavior, inducing them to exaggerate facial expressions or gestures to compensate for the poor quality of the interaction or the lack of other nonverbal cues (Fauville et al., 2021). Moreover, video calls typically restrict the visibility of an interaction partner to their face and upper body, and the video occupies only a small portion of our visual field. This is problematic not only because fewer cues are available, but also because, unlike during face-to-face interactions, we cannot rely on peripheral vision to detect gestures, shifts in posture, or subtle changes in eye gaze (Isaacs & Tang, 1994). Finally, genuine eye contact is not possible in a video call: you can either appear to make eye contact by looking at the camera, or receive it by watching the other person's eyes on the screen (Duffy & Benotsch, 2025). Nevertheless, people still feel the uncomfortable sensation of others constantly looking at them (Fauville et al., 2021). For all these reasons, although nonverbal cues are typically exchanged spontaneously and unconsciously, video calls demand greater cognitive effort and sustained attention to both express and interpret these cues (Fauville et al., 2021).

Interestingly, despite these constraints, automatic mimicry can occur in video call contexts (Gvirts et al., 2023; Marx et al., 2025). Even though the research on this topic is still limited, some studies suggest that being behaviorally aligned – even through video – may increase emotional contagion (Marx et al., 2025) and promote positive affect (Gvirts et al., 2023). These are promising results, but what remains unclear, however, is how mimicry in video-mediated interaction *compares* to that in face-to-face interaction, or to asynchronous, non-interactive video exposure. Studying mimicry in video calls is not only relevant from an applied perspective – given the growing importance of digital communication – but also offers a unique opportunity to advance our theoretical understanding of mimicry as a social mechanism. Directly comparing face-to-face interaction with digitally mediated contexts allows us to probe the role of real-time social contingency and interactional feedback in driving mimicry. Does the mere presence of visible cues suffice to trigger mimicry, or is dynamic, reciprocal interaction necessary? Moreover, how do these dynamics play out for behaviors that may have stronger adaptive significance in co-present interaction, like yawning and scratching? The present dissertation moved beyond the question of whether mimicry occurs in video contexts to ask how its strength and effects depend on the level of interactivity or the nature of target behavior.

Perception of Social Cues in Artificial Agents

As technological progress has reshaped how we interact with one another, it has also opened the door to an entirely different kind of social partners: artificial agents[1] . From healthcare robots to virtual tutors, these agents are being deployed in settings where social interaction plays a central role (Broadbent, 2017). Deploying artificial agents in the real world is not straightforward; it hinges on whether people recognize and respond to them as social beings. That is, whether we see them as social actors in the first place. Interestingly, people often do treat artificial agents as social partners, sometimes even when they are clearly non-human (Zhou et al., 2024). This phenomenon is explained by the media equation theory (Reeves & Nass, 1996), which proposes that people tend to treat media and machines in fundamentally social ways, often unconsciously. We greet voice assistants, attribute intentions to robots, and get angry with computers because the human brain developed to interpret certain cues as socially meaningful, regardless of their source.

However, "socialness" is not a black-and-white concept. Rather, it exists along a continuum influenced by factors such as the agent's appearance, its behavior, the social context, and the expectations and beliefs of the human perceiver (Hortensius & Cross, 2018). One important factor in this context is expressiveness – the variety of nonverbal cues the agent produces to signal emotion, intent, or

1 Based on Hortensius & Cross (2018), in this dissertation I use the term artificial agents to refer to physical robots (human- or machine-like), virtual agents (including virtual avatars or characters), and virtual assistants.

responsiveness. Research shows that more expressive agents are perceived as more engaging, likable, and human-like (Hortensius et al., 2018). That is, an agent that displays dynamic nonverbal cues like facial expressions, vocal prosody, or gestures is more likely to be perceived as more intentional or having a mind of its own (Wiese et al., 2017). Even more subtle cues, such as pupil dilation (Dong et al., 2022), face blushing (Pan et al., 2008), or gaze behavior (Fiore et al., 2013), can enhance expressivity and social presence. In this sense, expressiveness functions as a proxy for mind: when a non-human entity behaves in ways that suggest it can feel, think, or react appropriately in social contexts, observers are more inclined to treat it as a social and intentional partner (Waytz, Gray, et al., 2010).

Given the heterogeneity in artificial agent designs, spanning from virtual avatars to machine-like robots, the types of cues available to the agent depend greatly on its form and embodiment. Yet one element is consistently present across nearly all social agent designs: the eyes. In both human-human and human-agent interaction, eyes are a potent and rich social cue, essential for establishing mutual attention, conveying mental states, and regulating the flow of interaction (Admoni & Scassellati, 2017). Regardless of how a human-like artificial agent appears, eyes are almost always included in social agent design, precisely because of their central role in human social cognition. Designing biologically inspired eye behaviors in artificial agents – such as gaze shifts, blinking, or pupil dilation – heightens an agent's perceived expressivity and emotional salience (Dong et al., 2022; Fiore et al., 2013): eyes that move, blink, or track gaze convey social presence (Fiore et al., 2013), which can affect human behavior just like another human would. For instance, direct eye contact from a robot can make participants hesitate or revise their answers (Onnasch et al., 2023), while subtle changes in pupil size can foster a sense of familiarity and positive affect (Xue et al., 2022), even in non-humanlike agents. In some cases, such cues can even modulate relational dynamics in ways similar to face-to-face interaction: for example, direct gaze can promote honesty in self-report tasks, mirroring natural social mechanisms (Schellen et al., 2021). These findings illustrate how critical nonverbal eye cues are, not just for aesthetic realism, but for shaping cognitive, emotional, and behavioral responses in human observers.

A key question for the present dissertation is how behavioral responses to nonverbal cues, such as pupil dilation, depend on the nature of the agent displaying them. The choice to investigate pupil size over other cues has two main advantages: it is a subtle cue that most people perceive unconsciously, and it is easy to manipulate across different artificial agents and humans alike. We know that, in human-human interaction, changes in pupil diameter can signal arousal, interest, or surprise (Kret & Fischer, 2018), and can influence the perceiver's attitude and prosocial behavior (Prochazkova et al., 2018b; van Breen et al., 2018). But do these same prosocial effects emerge when this subtle cue is embedded in an artificial agent, and if so, does it matter what kind of agent conveys the cue? Answering this question is crucial not only to inform socially aware agents' design but also to advance our own understanding of how subtle cues function across different interaction contexts. Paraphrasing Broadbent (2017): artificial agents offer a mirror through which we can examine our own nature and better understand how we perceive the (social) world around us.

Cultural Modulation of Cue Interpretation and Technological Trust

Cues embedded in humans and agents can influence our behavior towards them, yet our responses also depend on our own idiosyncrasies. Human traits, such as personality, socioeconomic background, and culture, play a significant role in shaping our social interaction with both humans and machines. Culture, in particular, is central to the present dissertation, as prior studies observed cross-cultural variation in emotional perception (Kawahara et al., 2021), attitudes toward artificial entities (Nomura, 2017), and prosocial behavior (Hugh-Jones, 2016) – all core domains explored in the dissertation. A cross-cultural lens is vital in human–robot interaction research, given the wide fluctuation in how societies perceive, interact with, and emotionally respond to artificial agents (Lim et al., 2021). One contributing factor for this variety is that nonverbal cues are produced and perceived differently across cultures (Haensel et al., 2020; Marsh et al., 2003), also when they come from artificial faces (Yamamoto et al., 2020). For instance, individuals from East Asian cultures tend to focus on the eyes to perceive emotion, while those from Western cultures rely more on the mouth (Jack et al., 2009; Kawahara et al., 2021; Yuki et al., 2007). These perceptual differences are rarely considered in agent design, yet they could potentially affect human-agent interactions in culturally specific and unexpected ways.

One country frequently highlighted in cross-cultural human interaction research is Japan, given its long history of production and social integration of robots (Glikson & Woolley, 2020b). In Japan, people report warmer feelings (MacDorman et al., 2009) and more emotional engagement with artificial agents, which can be as profound as with humans (Karpus et al., 2025). This acceptance is thought to be influenced by Shinto beliefs attributing spiritual essence to inanimate objects, fostering openness to social and emotional integration of artificial agents (Jensen & Blok, 2013). Furthermore, given the significance of pupil size for some chapters of this dissertation, Japan represents an even more interesting case because it combines the well-known favorable view of artificial agents (MacDorman et al., 2009) with a culturally specific emphasis on eye-based emotional cues (Yuki et al., 2007). Japanese culture places particular importance on the eyes, especially pupil size, as critical emotional signals – a tendency observed in both empirical research (Haensel et al., 2020; Yamamoto et al., 2020) and popular media like manga and anime, where large pupils signal warmth and small pupils coldness. The cross-cultural element broadens the scope of the dissertation by considering how cultural context shapes the perception of nonverbal cues in artificial agents and whether this, in turn, differentially influences prosocial behavior across cultures. For the many differences in emotional perception and social acceptance of agents, Japan and Europe represent compelling cultural counterparts. If agents may serve as a mirror to human behavior, then cross-cultural comparison becomes essential for understanding which responses are rooted in shared human tendencies and which are shaped by culturally specific norms and expectations.

Outline of the Dissertation

The present dissertation investigates the role of nonverbal cues and their mimicry in shaping various forms of prosocial behavior across a wide range of contexts – from real-life human interactions to technology-mediated communication and encounters with artificial agents. While mimicry has traditionally been associated with affiliation and bonding, this dissertation critically evaluates such assumptions by exploring how the meaning and function of mimicry shift depending on the cue involved, the social context, and the interaction medium. While the early chapters (Chapter 2, 3, and 4) focus on automatic mimicry, the later chapters (Chapter 5 and 6) shift toward understanding how subtle social cues such as pupil size are interpreted depending on the nature of the social partner (human or artificial), and the cultural background of the perceiver. I will pursue these questions across six chapters, combining naturalistic studies and cross-cultural experimental designs to investigate how social cues and their mimicry function across diverse agents, contexts, and cultures.

In Chapter 2, I present a comprehensive theoretical review of automatic mimicry, encompassing both motor and autonomic domains. Drawing from research in both humans and non-human animals, the chapter critically evaluates dominant views that position the function of mimicry as affiliative and introduces alternative accounts that conceptualize mimicry as tools for social and environmental prediction. The chapter emphasizes the importance of considering the nature of the cue being mimicked, the valence of the interaction, and the contextual demands that shape the function of mimicry. While this theoretical framework opens the dissertation, it was developed in response to and informed by the empirical findings in later chapters.

In Chapter 3, I investigate how the visibility of social cues impacts motor and autonomic mimicry, and whether this, in turn, shapes trust and reciprocity. In a naturalistic dyadic set-up, participants played several trials of a trust game in a condition where they could see each other and where they could not see each other. This design allowed me to probe the emergence and effects of autonomic mimicry by examining what happens when visual information is removed from the interaction. Specifically, I explored I) whether and how autonomic mimicry predicts prosocial choices during a trust game, II) whether visual access to social cues affects the emergence and effects of synchrony, and III) whether autonomic and motor mimicry co-occur in real-time social exchanges. The overall goal was to understand the importance of visible cues and contextual affordances in determining the functional role of mimicry.

Given the importance of visual access not only for mimicry to emerge but also to influence trust choices, as shown in Chapter 3, in Chapter 4 I investigated to what extent visual access needs to be available for mimicry to arise and function effectively. Moving from human-human interaction to human-mediated interaction, I examined whether real-time contingency is necessary by comparing three interaction settings: face-to-face interaction, video calls, and pre-recorded video. This allowed me to tease apart the role of temporal contingency and visual feedback in shaping mimicry and its effects on social outcomes like trust. Going beyond previous literature, I focused on less commonly studied behaviors, scratching and yawning, that we share with many non-human species, are highly contagious, and whose adaptive significance may be particularly pronounced in co-present, face-to-face interaction. These behaviors offer a unique lens through which to observe mimicry's modulation by digital environments.

Chapter 5 starts the transition from interaction between humans to human-robot interaction by introducing a cross-cultural perspective on the perception of artificial agents. In anticipation of Chapter 6, this chapter aims to first establish whether differences exist in how artificial agents are perceived depending on their shape (human-like vs. machine-like) and on the cultural background of the perceiver. Specifically, I investigated whether Dutch and Japanese participants differ not only in their explicit attitudes toward human and artificial agents, but also in their more automatic, implicit responses, two levels that rarely co-existed in prior research. By examining preferences for different degrees of human-likeness and agent shape, Chapter 5 lays the empirical groundwork for exploring how cultural background shapes the perception and social meaning of subtle cues in humans and artificial agents.

Building directly on this, Chapter 6 investigates how a subtle nonverbal cue – pupil size – modulates prosocial behavior, specifically honesty, in real interactions with humans, robots, and avatars across cultures. By manipulating pupil size across all agents and testing a large heterogeneous sample of European and Japanese participants, the study asks whether the same cue may evoke different behavioral responses depending on the cultural background and the embodiment of the agent. Given that pupil size is a particularly easy parameter to manipulate experimentally, and that prior work linking pupil size to increased liking and prosocial behavior is largely correlational, we also aimed to experimentally test the causal effects of pupil size on social outcomes. The agent types used in this study were informed by the findings of Chapter 5, ensuring that neither avatar nor robot representations were biased by pre-existing shape preferences. The results of Chapter 6 offer critical insight into how subtle social signals are perceived cross-culturally and how agent embodiment interacts with cue perception, laying the foundation for future research on the mimicry of such cues in human–agent interaction.

Finally, Chapter 7 zooms out to speculate on the future of social robotics. Rather than focusing solely on human-robot interaction, it advocates for a "bottom-up" approach that uses robot-robot interaction to model simple social behaviors. Drawing inspiration from evolutionary robotics and biologically inspired mechanisms such as hormonal modulation, the chapter argues for a new framework to study foundational principles of sociality in artificial systems.

Together, these chapters argue for a more flexible, context-sensitive understanding of mimicry and nonverbal cues – one that accounts for the interaction setting, the social partner, and the cultural lens through which cues are perceived. In doing so, the dissertation contributes not only to the fundamental understanding of how automatic mimicry works but also to the design of artificial systems capable of navigating human social environments in more nuanced and culturally attuned ways.