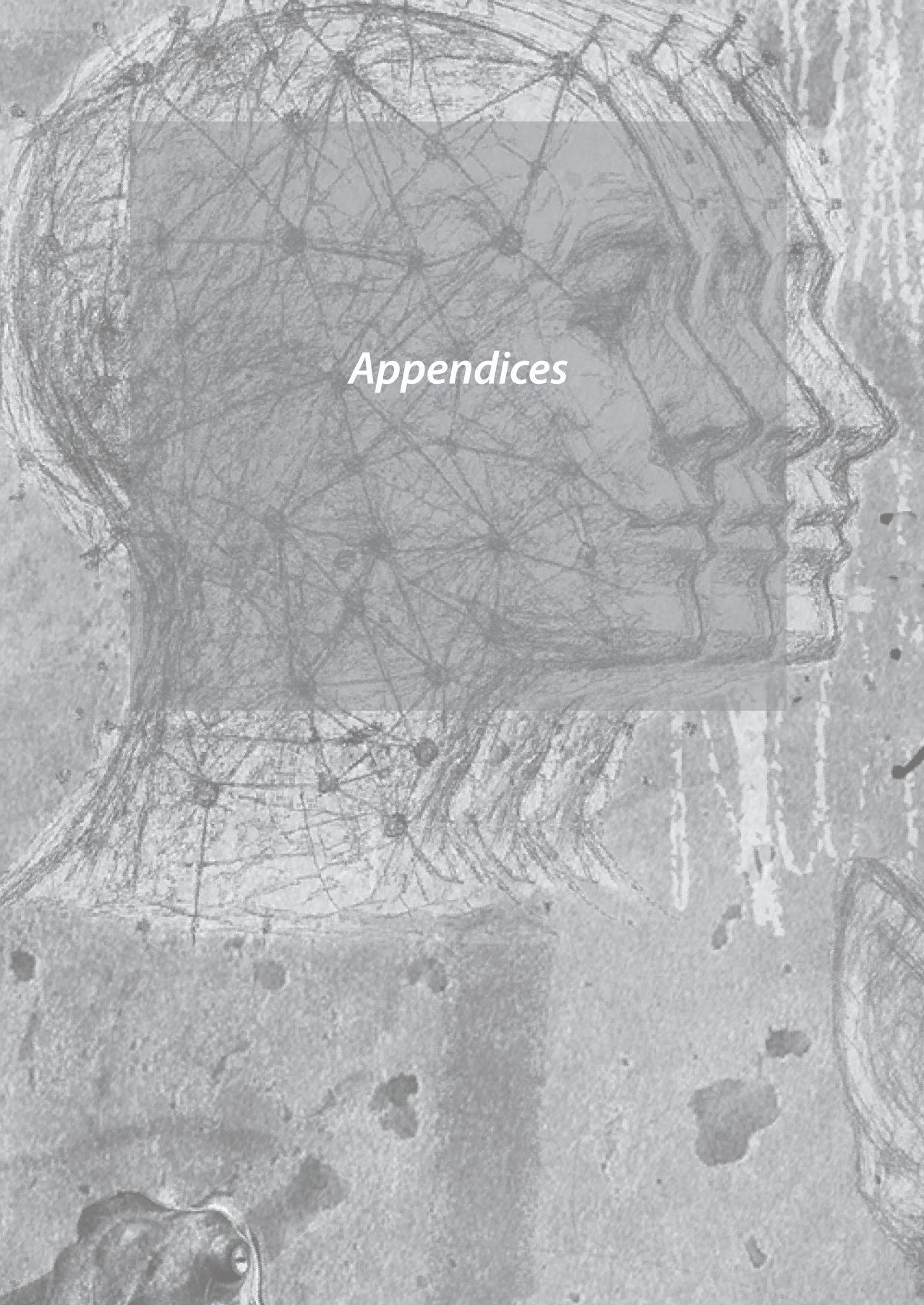
Interactional beings: the power of automatic mimicry and nonverbal cues in shaping human-human and human-robot naturalistic interactions

Diana, F.



Appendices

Summary

Life depends on interaction. From ants coordinating through pheromones to whales synchronizing movements and primates grooming for social bonding, cooperation, and alignment are vital for survival. Humans, too, are deeply social. Much of our social navigation relies on nonverbal cues such as facial expressions, body movements, vocal prosody, and physiological signals. These cues give us fast, often unconscious access to what others feel and intend. Their power lies not only in what they communicate, but also in the fact that we often copy them. By mirroring each other's behavior, people become aligned.

This phenomenon is known as automatic mimicry. It refers to the unintentional copying of another person's actions or bodily states and can be observed across many species. Mimicry includes visible behaviors such as smiling, yawning, or gesturing, but also less visible processes such as matching another person's heart rate, pupil size, or skin conductance. For a long time, mimicry has been described as a kind of "social glue": a mechanism that promotes liking, empathy, cooperation, and social bonding. However, this view does not fully capture how mimicry works. People also mimic negative or ambiguous behaviors, such as anger, fear, nervous scratching, or yawning, and these do not consistently lead to positive social outcomes. Instead, the effects of mimicry depend on what is being mimicked and in which situation. Mimicry can smooth social interactions, help regulate tension, or even strengthen avoidance and discomfort. Rather than being uniformly prosocial, mimicry appears to be a flexible process that supports alignment in different ways depending on context. This flexibility becomes especially important in modern forms of interaction that rely on technology. Video calls, for example, allow people to see and hear each other, but they also distort or remove important cues through delays, limited visibility, poor resolution, or the lack of true eye contact. Even under these constraints, mimicry can still occur. What remains unclear is whether mimicry in mediated interactions is as strong or as meaningful as in face-to-face encounters, and whether real-time back-and-forth feedback is necessary for mimicry to influence behavior. Studying mimicry in mediated settings is therefore essential for understanding how timing, feedback, and contingency shape social alignment.

Similar questions arise with the increasing presence of artificial agents such as avatars, chatbots, and robots. People often react to these systems as if they were social partners, relying on the same automatic processes they use in interactions

with other humans. Expressive behavior plays a key role in this. Artificial agents that display dynamic nonverbal cues, especially in the eyes, are more likely to be seen as intentional, engaging, and socially meaningful. Subtle cues such as gaze direction or pupil size can influence trust, honesty, and approach behavior, even when they are generated by machines. Yet it is unclear whether such effects generalize across different types of artificial agents or across cultural contexts.

This brings culture into focus. Research shows that people from different cultural backgrounds attend to different facial regions when interpreting emotions, and that societies vary in their familiarity with and acceptance of artificial agents. In Japan, for example, people place high cultural value on eye-based cues and have a long history of integrating robots into daily life. More broadly, culture shapes all the main themes of this dissertation. It influences how people mimic others, how they perceive and express social cues, how they evaluate dishonest behavior, and how they respond to artificial agents. Cross-cultural research is therefore crucial for separating responses that are broadly shared from those that are culture-specific. Together, these considerations motivate the central goal of this dissertation: to examine whether and how social cues and their mimicry shape prosocial behavior, and how these effects vary across interaction contexts, such as face-to-face, video-mediated, and human–agent interaction, and across cultures in Europe and Japan.

After setting up the theoretical basis of the thesis in Chapter 1, Chapter 2 sets out a theoretical framework for understanding mimicry. This chapter challenges the dominant assumption that mimicry primarily exists to promote affiliation. Instead, it proposes that mimicry also serves a predictive function: it helps individuals anticipate others' behavior and, by extension, the environment. The chapter begins with a review of the traditional social glue hypothesis, which links mimicry to trust, bonding, and group cohesion. Although many studies support this link, their results are often context-dependent, inconsistent, or contradictory. The chapter then reviews evidence showing that mimicry can serve other purposes, such as monitoring potential threats or navigating uncertain social situations. Focusing only on positive, affiliative outcomes risks oversimplifying the phenomenon. Negative emotions and ambiguous signals can be adaptive, and excluding them limits our understanding of what mimicry does. The chapter, therefore, develops three core arguments: first, mimicry is not limited to positive expressions; second, mimicry does not always depend on an intention to affiliate;

and third, any affiliative effects may follow from more immediate benefits to the person who mimics. Building on these points, the chapter introduces a predictive model of mimicry, in which copying others helps individuals better anticipate events and respond quickly. Mimicking another person's gaze or movement, for example, may improve detection of danger or changes in the environment. The chapter also outlines possible mechanisms and evolutionary pressures that could have shaped mimicry as a predictive tool, and proposes ways to test this idea experimentally. The aim is not to reject affiliation as a function of mimicry, but to complement it with prediction as an additional explanation. Mimicry can help people gather information and navigate social and physical environments more effectively. In this view, mimicry is not a single-purpose social tool, but a context-sensitive mechanism that supports prediction and adaptation.

In Chapter 3, I directly examined mimicry across different levels to test whether it serves a prosocial function and to explore the mechanisms by which it emerges. Using a naturalistic dyadic set-up, participants played multiple rounds of a trust game under two conditions: one where they could see each other and one where they could not see each other. This design allowed me to probe the emergence and effects of autonomic mimicry, the alignment of physiological responses, by assessing what happens when visual information is removed. The study addressed three main questions: whether autonomic mimicry predicts prosocial decisions, whether visual access influences the emergence and impact of synchrony, and whether physiological and motor mimicry unfold together during interaction. The overarching aim was to understand how visible cues and contextual affordances influence the functional role of mimicry. From a functional perspective, I found that autonomic mimicry was not linked to greater prosociality; instead, higher mimicry predicted reduced reciprocity. This challenges the notion that mimicry always reflects a drive to affiliate or cooperate. From a mechanistic perspective, the findings shed light on how mimicry operates. Autonomic mimicry emerged in both conditions, but it was significantly stronger when participants could see each other. More importantly, mimicry influenced behavior only with visual access. Although physiological synchrony emerged in both conditions, it was stronger when participants could see each other, and it influenced behavior only when visual access was available. This suggests that while synchrony can arise without sight, interactive visual contexts are necessary for it to carry social meaning.

Interestingly, increased facial expressivity did not explain physiological coupling, pointing to the role of more subtle cues, such as pupil size.

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 how much visual access is needed for mimicry to occur and influence behavior. The focus shifts from face-to-face interaction to mediated interaction by comparing three settings: in-person interaction, live video calls, and prerecorded video. This design separates the role of real-time feedback from physical co-presence. The chapter focuses on behaviors such as scratching and yawning, which are common, contagious, and shared across species. These behaviors provide a useful window into how mimicry operates under different technological constraints. The results show that mimicry was strongest when real-time visual feedback was available, both in face-to-face and video-call interactions. Physical presence was not required. What mattered was temporal contingency. Mimicry was disrupted when participants watched prerecorded videos. Scratching was mimicked most frequently, followed by face-touching and yawning, suggesting that frequent, low-effort behaviors are particularly sensitive to alignment. Mimicry showed little relationship to empathy, anxiety, or trust, reinforcing the idea that it may reflect interactional alignment rather than deliberate prosocial intent. Together, the findings show that real-time visual contingency is sufficient for mimicry and trust-related processes to emerge, whereas delayed or non-contingent input undermines both.

Chapter 5 starts the transition from interaction between humans to human-robot interaction by introducing a cross-cultural perspective on how artificial agents are perceived. 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. At the implicit level, participants from both cultures showed a small but consistent preference for humans over artificial agents, regardless of agent shape. Cultural differences emerged mainly in explicit reports, and explicit and implicit attitudes were only aligned in the Dutch sample. Neither robot nor avatar form elicited strong implicit differentiation, suggesting that static images are not sufficient to trigger deep automatic responses. These findings show that cultural differences in attitudes toward artificial agents are more visible in conscious evaluations,

while implicit preferences remain relatively stable. Chapter 5 therefore provides a foundation for studying how culture shapes the interpretation of subtle social cues in interactions with artificial agents.

Chapter 6 builds on this foundation by examining how a single subtle cue, pupil size, affects honesty in interactions with humans, robots, and avatars across cultures. Pupil size was experimentally manipulated across all agent types in a large and heterogeneous sample of European and Japanese participants. The study tested whether the same cue produces different behavioral effects depending on culture and embodiment. 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 agents used were selected based on the findings of Chapter 5 to avoid strong pre-existing preferences. The results reveal clear cultural and embodiment-specific patterns. European participants were more honest with human partners than with robots or avatars, while Japanese participants showed no such differentiation. Pupil size influenced dishonest behavior only in interactions with avatars, and in opposite directions across cultures: Europeans lied more to avatars with large pupils, whereas Japanese participants lied more to avatars with small pupils. No pupil-size effects emerged for human or robot partners. Physiological and gaze data confirmed that participants processed the agents differently, but these measures did not explain the cultural differences in behavior. These findings show that subtle cues do not have universal meanings. Instead, their behavioral impact depends on how the cue is interpreted within a specific agent category, highlighting the importance of perceptual ambiguity and cultural priors in human–agent interaction.

Finally, Chapter 7 broadens the perspective to consider the future of social robotics. Rather than focusing exclusively on human–robot interaction, the chapter argues for a bottom-up approach that studies robot–robot interaction as a way to model basic social processes. Drawing on evolutionary robotics and biologically inspired mechanisms such as hormonal modulation, the chapter proposes that social behavior can emerge from simple interaction rules rather than being explicitly programmed. This approach aligns with the dissertation's broader argument that alignment, coordination, and sensitivity to cues arise from interaction dynamics.

Studying these processes in artificial systems may reveal general principles of social behavior shared across biological and artificial agents.

In sum, this dissertation examined how nonverbal cues and their mimicry influence social behavior across contexts, agents, and cultures. Using naturalistic dyadic studies, mediated interactions, cross-cultural experiments, and human–agent designs, it challenges the assumption that mimicry is inherently affiliative. Instead, mimicry is shown to be context-dependent, shaped by the type of cue, the interaction setting, and the nature of the partner. The findings support a view of mimicry as a flexible mechanism that helps individuals navigate and predict their social environment. The work also demonstrates that subtle cues such as pupil size can carry different meanings across cultures and embodiments. Methodologically, the dissertation emphasizes the value of studying real interactions, even at the cost of some experimental control. Conceptually, it calls for broader theories of mimicry that incorporate prediction, negative outcomes, and multiple levels of alignment. Practically, the findings highlight the importance of culturally sensitive design in social robotics, showing that social cues do not operate universally. Ultimately, I hope this dissertation meaningfully contributes to the literature on automatic mimicry and nonverbal cues across the diverse interaction contexts and partners that our ever-evolving social world has to offer.