SENSORIMOTOR COMMUNICATION IN JOINT ACTION: THE ROLE OF DISTAL GOALS

Martin Dockendorff

In partial fulfilment of the requirements for the degree of Doctor of Philosophy in Cognitive Science

Primary supervisor: Günther Knoblich

Secondary supervisor: Cordula Vesper

Doctoral advisor: Laura Schmitz

Vienna, Austria

2024

COPYRIGHT NOTICE

Declaration of authorship

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or which have been accepted for the award of any other degree or diploma at Central European University or any other educational institution, except where due acknowledgment is made in the form of bibliographical reference.

The present thesis includes work that appears in the following publications:

Dockendorff, M., Sebanz, N., & Knoblich, G. (2019). Deviations from optimality should be an integral part of a working definition of SMC: Comment on" The body talks: Sensorimotor communication and its brain and kinematic signatures" by Pezzulo et al. *Physics of life reviews*, *28*, 22-23.

Dockendorff, M., Schmitz, L., Vesper, C., & Knoblich, G. (2023). Understanding others' distal goals from proximal communicative actions. *Plos one*, *18*(1), e0280265.

Dockendorff M, Schmitz L, Vesper C, Knoblich G (2024) Communicative modulations of early action components support the prediction of distal goals. *Plos one*, 19(6): e0306072.

Other parts of the dissertation will be submitted for publication with the following co-authors:

Chapter 4 with Laura Schmitz, Günther Knoblich and Cordula Vesper.

Martin Dockendorff

Abstract

The present thesis aims at investigating a form of communication whereby people systematically modulate the kinematics of the movements they carry out as part of a joint action. These kinematic modulations have the effect of making the actor's goals easier to predict, thereby facilitating interpersonal coordination. So far, this form of "sensorimotor communication" has mostly been studied in contexts in which co-actors modulate their actions to inform each other about their immediate, *proximal* goals (e.g., reaching for a cup). This thesis expands this focus by asking whether people can also rely on modulations that convey information about *distal* goals (e.g., reaching for a cup in order to pass it to a friend). Thus, while proximal goals are directly tied to the action leading to their achievement, distal goals are both spatially and temporally separated from the action by one or more intermediate actions, which together constitute an action sequence. The first study presented starts by looking at whether observers interpret communicative modulations of simple sliding movements to predict another person's distal goal. The findings show that observers rely on modulations in movement velocity, enabling them to predict both proximal and distal goals. A second study further explores *how* observers make these predictions about distal goals, particularly when the communicative modulations occur in the first step of a two-step action sequence. The findings not only replicate those from the first study, but also indicate that observers can often make these predictions while only relying on modulations present in the first step of the sequence. The third study moves from the observation of communicative modulations to their production in an interactive context, and asks whether co-actors are able to solve a coordination problem by means of communicative modulations. The findings indicate that co-actors can rely on each other's communicative modulations to solve this problem, leading them to establish a communication system based on these modulations. Overall, the findings presented in this thesis suggest that people can both interpret and actively inform others about goals that expand from the here-and-now towards the (predictable) future.

Acknowledgments

The first person I would like to thank is Günther Knoblich. Günther was my PhD supervisor, but with the passing of time has become a mentor of sorts. He supported me in many different ways and trusted me by letting me get away with my regular distractions. But maybe more importantly, he became a constant reminder that the best way of approaching things, especially in academia, is with a certain degree of lightness and, of course, humor.

Second, Cordula Vesper and Laura Schmitz. Their contribution to this thesis was essential. Before starting the PhD I read a paper by Cordula that had a big influence on my way of thinking about communication. I hope some of that influence made its way into this thesis. On top of that, she was a source of constant support, ideas, advice and much-needed solutions, at all times. Laura bears the responsibility of having introduced me to the research on "joint action" back in 2017, and since then she's been a steady colleague and friend, while also being the most thorough reader and critic of my work. For all this, and much more, thanks to both of you.

Other people who were important in these last couple of years include: Barbu Revencu, Elena Marx, Arianna Curioni, Luke McEllin, Laura Schlingloff-Nemecz, Nima Mussavifard, Thomas Wolf, Nejra Rizvanović, Maximilian Marschner, Vjeran Kerić, Bálint Varga, Oliwia Iwan, Francesca Bonalumi and Ieva Lukošiūnaitė.

Finally, for making things run smoothly in the background, I want to thank Réka, Andi, Fanni, Candasch, Eszter and Dávid.

Table of Contents

Chapter 1. Introduction

Many of the things we do in our day-to-day life we do with others. From having a conversation, giving a handshake or passing a cup to someone, to playing a piano duet or moving a large piece of furniture – all of these involve, either by necessity or by choice, people doing something *together* (Clark, 1996; Levinson & Enfield, 2020). In order to succeed at performing such a wide variety of activities, people need to adapt their behavior to what others are doing: when talking, speakers need to adjust the timing of their utterances to avoid constantly interrupting each other, while at the same time minimizing the silences between turns (Sacks et al., 1974; Stivers et al., 2009); when passing a cup, the "giver" has to plan her movements in ways that take into account whether the "receiver" is ready to accept the cup and what she plans to do next (e.g., drink from the cup or place it on a table (e.g., drink from the cup or place it on a table; D. Rosenbaum et al., 1990); when playing a duet, pianists need not only to synchronize their entrances and exits, but also pay attention to how loud they play forte and piano, among other things (Goebl & Palmer, 2009). All these examples are forms of *a joint action*, in which two or more people coordinate their behavior in space and/or time to achieve a joint goal (Sebanz et al., 2006), where such goals include anything from achieving mutual understanding in conversation to cooking a paella with a friend (Knoblich et al., 2011; Sebanz & Knoblich, 2021; Vesper, Abramova, et al., 2017).

What are the cognitive mechanisms that enable people to engage in joint actions? Is there a single overarching mechanism that makes most joint actions possible, or multiple ones contributing to different forms of joint action? And finally, can we address these questions empirically, through psychological experiments?

One aspect in which research in joint action has been especially active is in its effort to go beyond the use of traditional psychological methods to study social interaction (Schilbach et al., 2013). These traditional methods often involve asking participants to look at pictures of socially relevant stimuli presented on screens (e.g., a face expressing a negative emotion, point-light displays of a moving body) or read vignettes describing the actions and beliefs of one or more actors, and to react to them in ways specified in advance by the experimenter (Roepstorff & Frith, 2004). Because of their simplicity, these methods have provided researchers with well-controlled experimental designs and replicable methods. They have also contributed important findings relating to how we process social stimuli. However, this same simplicity comes with an obvious limitation, which is that they fail to capture the more complex and richer dynamics of people's behavior in real-time interactions, particularly those in which we continuously adapt our actions to what others are doing (as well as to what *we think* they are doing), and vice versa. Capturing this form of mutual adaptability, which is at the core of most joint actions, has been one of the main challenges of researchers working in this area (Becchio et al., 2010; De Jaegher et al., 2010; Schilbach et al., 2013).

This thesis is concerned with one of the most well-studied forms of joint action: communication (Bavelas, 2022; Clark, 1996; Finnegan, 2013). More specifically, with a form of nonverbal communication ("sensorimotor communication") to which I will turn later in the Introduction. For now, I will look at communication from a broader perspective, which is precisely the perspective that many who have studied it as joint action have adopted.

Looking at communication as a joint action immediately raises two related questions. First, in what sense is communication a joint action? And if it is a joint action then what, if anything, gets coordinated when we communicate? The answer to these questions depends very much on the specific aspect of communication that is being considered (e.g., meaning, turn-taking, modality), and more broadly on the theoretical approach adopted. Yet, what seems to be clear is that communication, at least in its most frequent forms, involves two or more people doing something together, whether it is having a casual conversation or discussing what to cook for dinner (Clark, 1996). But while having a conversation or discussing what to cook are both clear cases in which people make use of communication, they differ in one crucial way. This difference, I will argue in what follows, represents two different ways of approaching communication from a joint action perspective (Bangerter & Clark, 2003).

The first approach involves seeing communication as a form of joint action *in and of itself*. Here, speakers coordinate their communicative acts (e.g., spoken words, gestures, facial expressions, etc.) to achieve goals that are specific to communication, chief among which is reaching mutual understanding (Grice, 1989; Clark, 1996). Most casual conversations can be seen from this perspective, since in most of them speakers are simply required to keep the conversation going for a given amount of time (see "phatic communication", Malinowski, 1949). Keeping a conversation going appears, on the face of it, like an easy thing to do. However, this apparent ease is made possible because speakers keep close track of what is being said, by whom, when, to whom, and so on (Clark, 1996). In other words, even in the most mundane of conversations, speakers need to coordinate on several aspects that are specific to communication, from the topics covered (e.g., "are we talking about the same thing?") and the words used (e.g., "are we using the same name?") to how they use these words and when. All this, and possibly more, is needed for understanding each other. Different theoretical frameworks have been built around the idea that communication is itself a form of joint action, but most of them have focused on a particular modality and context of communication, namely language use in natural conversation, i.e., dialogue (Garrod & Pickering, 2009).

The second way of approaching communication from a joint action perspective highlights the role that communication plays as a means through which people facilitate the achievement of other, possibly instrumental joint goals, like cooking dinner together or assembling a TV stand. In these cases, communication can still be seen as a joint action. However, the focus is on its role in facilitating the achievement of goals that go beyond merely understanding each other. Thus, discussing what to cook is a clear form of communication, but it can also be seen as facilitating the achievement of another joint goal, namely cooking something together later in the evening. Here, communication fulfills a coordination function: it helps people carry out instrumental joint actions by, for example, distributing their individual roles and corresponding actions (e.g., who will do the groceries, who will do the dishes), as well as their timing and location (Clark, 2005). A general term to refer to this coordination function of communication is that of a *coordination smoother* (Vesper et al., 2010).

In what follows, I will first review two ways of approaching communication from the first perspective I introduced, where it is seen as a joint action in and of itself (1.1 Communication as joint action). Despite the differences between these two approaches, they both share the assumption that communication involves some form of coordination, either in the sense of people following general principles by which they manage to communicate successfully (1.1.1 Principles that make communication a joint action), or in the sense of sharing some aspects of their mental representations while they interact (1.1.2 Mechanisms that make communication a joint action). Then, I will turn to the second perspective, that of communication as a "coordination smoother" (1.2 Communication as coordination smoother), and zoom in on sensorimotor communication as an important coordination smoother (1.2.1 Sensorimotor communication), Finally, I will raise open questions in research on sensorimotor communication that will serve as a basis for my subsequent, empirical chapters (1.3 Questions addressed in this thesis).

1.1 Communication as a joint action

1.1.1 Principles that make communication a joint action

Let's begin with a trivial observation: when speakers communicate, they produce what we can broadly call "communicative acts"" – including spoken and/or signed words/utterances, gestures, facial expressions and so on – with the primary goal to be understood by a receiver. Understanding communicative acts, however, is far from obvious, as shown by how often speakers and receivers need to use strategies to resolve troubles in hearing and understanding (e.g., in dialogue, the use of *huh?*, *what?* ; E. A. Schegloff, 1991). Often, such difficulties arise because communicative acts, and especially linguistic ones, can only be properly understood if receivers take into account not only what a speaker is explicitly saying or showing with a given act, but what he *intends* to convey with it (Grice, 1957a). To take a textbook example: the utterance "It's hot in here", taken literally, simply describes the perceived temperature of a room; however, its *intended* meaning is often taken as an indirect request directed at a receiver so that she opens a window or turns on the air conditioner (Gibbs, 1983; Brown & Levinson, 1987). This simple example illustrates a common (some would even say pervasive, e.g., T. Scott-Phillips, 2014) feature of human communication, involving a clear disparity between what is explicitly communicated (i.e., what is *said*) and what is meant (i.e., what the speaker *means* with what she's saying) – technically known as underdetermination (T. Scott-Phillips, 2014). Indeed, underdetermination has been argued to be such a common feature of human communication that it can even be found in simple non-linguistic acts, like showing or pointing, where one also finds disparities between what speakers are showing or pointing at and what they intend to convey with such an act (Sperber & Wilson, 2015; Wharton, 2009). These observations raise an interesting question: if underdetermination is so present in human communication, and particularly in language, then how is it that we are so good at quickly understanding others and making ourselves understood?

According to an influential philosophical tradition, initiated by the British philosopher Paul Grice, speakers can bridge the disparity between what they say and what they mean by adhering to certain "cooperative principles" when designing their communicative acts. By following these principles, speakers make explicit what they mean, and this then helps receivers interpret and understand what speakers are trying to communicate (Grice, 1957b, 1989). A detailed discussion of these different principles is beyond the limits of the present introduction (Grice, 1989; Sperber & Wilson, 1995; Levinson, 2000; Noveck, 2018). For the present purpose, it is sufficient to say that according to this account communication is a joint action because it requires both speakers and receivers to follow one or several of these cooperative principles in order to reach mutual understanding. Speakers, on the one hand, adjust their communicative acts to these principles, thereby making their acts easier to understand; receivers, on the other, adjust their interpretation of these same acts by relying on the well-founded assumption that speakers are, indeed, following these cooperative principles when designing their communicative acts (Tomasello, 2008).

Following this idea, a proposal that has clearly spelled out a well-defined principle that governs communication, and in so doing identified what makes communication a form of joint action, is Herbert Clark's proposal of a principle of "least collaborative effort" (Clark & Wilkes-Gibbs, 1986; Clark, 1996). In a nutshell, this principle states that speakers will design their communicative acts in ways that maximize the effects they wish to produce in receivers, while at the same time minimizing the effort (in terms of length of words in the case of spoken language, or submovements in the case of signed language) that they themselves and their partners, as a dyad, need to invest in order to reach mutual understanding (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986). Part of the appeal of this principle is that it provides a concrete way of measuring the "jointness" of communication, specifically by looking at how various measures of effort are distributed *across* participants in the conversation (rather than *within* participants, see Foster-Cohen, 2004). Furthermore, this principle is general enough to capture the way in which people perform many other joint actions, from handing an object (D. Rosenbaum et al., 1990) to holding a door open for someone (Santamaria & Rosenbaum, 2011). In these cases, people also try to minimize their "joint" effort by, for instance, choosing to perform movements that are shorter for both themselves *and* their co-actors, rather than movements that are shorter for either themselves or their co-actors only (Török et al., 2019; see also Rasenberg et al., 2022). What this suggests is that the principle of least collaborative effort, initially proposed by Clark to describe how people distribute effort in communication, might turn out to be an instance of a more general principle, one of shared-effort or coefficiency (Török et al., 2019), that guides how people perform joint actions more generally.

One communicative situation in which the principle of least collaborative effort seems to play a key role is when speakers detect and address troubles during natural conversation, commonly known as "repair" (E. A. Schegloff et al., 1977; Dingemanse & Enfield, 2024). In natural conversations, a speaker who fails to understand what the other speaker just said, can initiate a repair sequence by asking a question (e.g., "the what?", "who?") use interjections (e.g., "huh?") or simply repeat what was said with an upward intonation (e.g., "You said we'll meet at *ten*?"). Then, the other speaker can provide a repair solution in the shape of an answer to the speaker's question or a clarification of what was previously said. Since repair sequences involve contributions from both speakers, and these contributions are mutually adjusted to address the trouble, they offer an ideal setting to look for evidence that speakers follow the principle of least collaborative effort in conversation.

To see this principle at work here, we first need to take a closer look at the possible actions that each speaker can choose from throughout the repair sequence. On the one hand, a speaker initiating a repair sequence by asking a question has the possibility of doing so with different levels of specificity: from open requests that give no specific information about what she failed to understand (e.g., "huh?", "what?"), or restricted requests about some particular element in the sentence (e.g., "who?"), to restricted offers that explicitly state the source of the misunderstanding (e.g., "the dog?") (Dingemanse & Enfield, 2015). On the other hand, a speaker who ends the repair sequence by clarifying what was misunderstood can also do it with varying levels of specificity (and, importantly, with a varying number of words), either because she repeats the whole sentence or because she repeats only the specific word that was misunderstood. Despite these varying levels of specificity, recent studies have shown that the way speakers collaborate to repair the trouble follows a logic that is consistent with the principle of least collaborative effort. Specifically, if the repair initiation of the first speaker is more specific, and thus indicates the source of the trouble, then less effort needs to be invested by the second speaker in repairing it (e.g., in terms of word length; Dingemanse et al., 2015). Similar findings have been reported in repair sequences that include gestures, where more gesture sub-movements are used the more specific the repair initiation, while fewer are used in its solution (Rasenberg et al., 2022). Thus, repair sequences can be seen as micro-level coordination problems arising in a larger joint action (i.e., the conversation) in which both speakers end up sharing the effort they invest in addressing the troubles they face throughout, leading ultimately to efficient and smooth conversations (Dingemanse et al., 2015; Rasenberg et al., 2022; Jara-Ettinger & Rubio-Fernandez, 2022).

Another, similar approach to the idea of communication as joint action focuses on the different strategies used by speakers to make sure that they have reached mutual understanding (Clark, 1996). This approach, also advocated by Herbert Clark, starts from the general observation that when we talk to each other we often rely on signals that fade rapidly (i.e, spoken or signed words), leaving no physical traces that could potentially help speakers remember the things they have said so far. Furthermore, most conversations unfold quickly and speakers are expected to contribute to them in a timely and "incremental" manner, that is, one step at a time, without constant repetitions, and in a way that builds on what was previously said (e.g., "tomorrow it will rain") to continue with what will come next (e.g., "you should bring an umbrella"). According to Clark, speakers achieve all this by relying on each other continuously during their conversations in a collaborative process called *grounding* (Clark & Brennan, 1991).

To illustrate how grounding works in conversation, we can look at simple exchanges in which speakers agree on the names they give to referents (Clark & Wilkes-Gibbs, 1986). Early studies by Clark and his colleagues studied these situations closely by asking pairs of participants to collaborate in a "referential communication task" (Krauss & Weinheimer, 1964). In these tasks, one of the participants (the director) is instructed to help the other one (the matcher) arrange twelve geometric shapes in a particular order, as fast and as accurately as possible. When looking at the transcripts of the conversations, researchers found that during the first couple of trials, directors were more likely to provide matchers with long descriptions of the shapes (e.g., "looks like a person who's ice skating"). But as the trials went on, directors gradually shortened the descriptions after each iteration (e.g., "the ice skater"), which then led to more efficient (i.e., overall faster) exchanges. According to Clark & Wilkes-Gibbs, descriptions become shorter because speakers collaborate by making each a contribution, one after the other. First, directors present a description or a name to matchers (i.e., the presentation phase). Then, matchers either accept or reject the description or name presented by providing evidence to directors that they have understood (or misunderstood) (i.e., the acceptance phase). This last phase can be done by means of backchannels (e.g., *uh huh* or *yeah*; E. Schegloff, 1982) or, in case of misunderstanding, by initiating a repair sequence. In making these sequential contributions, speakers end up grounding, or building up common ground, in conversation (Clark & Brennan, 1991; Brennan & Clark, 1996).

Grounding is, according to this perspective, an essential feature of human communication, found in practically any communicative situation in which speakers need to make sure that they are being understood by listeners. More generally, since this process requires both speakers to make their contributions in an orderly manner and in a way that takes into account each other's contribution, it clearly requires coordination of communicative actions. This coordination is key to achieve a specifically communicative joint goal, namely the goal of reaching mutual understanding. Grounding thus provides another clear illustration of the perspective that views communication as a form of joint action.

In sum, these previous approaches have been useful in highlighting the "jointness" of communication by showing how the production of even the simplest communicative acts, like naming an object, are better understood if we see them as part of a collaborative process, that is, as joint actions. These approaches are part of a long tradition, called the "language-as-action" tradition (Clark, 1996), whose focus has been mainly on describing how natural conversation is guided by people's explicit goals and intentions, including the goal of understanding each other (Garrod & Pickering, 2004). This more descriptive focus has often led these approaches to overlook the role played by low-level psychological mechanisms that presumably also contribute to communication. Recent proposals have specified such mechanisms and, in so doing, offered yet another way of looking at communication as a form of joint action. I now turn to one such approach.

1.1.2 Mechanisms that make communication a joint action

There are many reasons why having a conversation should be challenging for interlocutors. Besides the reasons I discussed in the previous section having to do with how speakers keep track of what is being said, there is also the challenge of being able to process ill-formed, or plainly ungrammatical sentences by speakers whose pronunciation and meanings, despite sharing the same language, differ greatly (e.g., a Viennese speaking to a Berliner). According to Garrod & Pickering (2004), what enables speakers to surmount these challenges is the fact that they come to *align* their mental representations of the situation at hand, particularly their linguistic representations. This form of alignment is automatic, since it relies on a simple mechanism whereby the production of a certain communicative act by one speaker, a word or a grammatical construction, primes the same representation in the mind of the other speaker, thus leading to similar communicative acts in both (i.e., repetitions). Since this convergence of mental representations and behavior across speakers is the direct result of their interaction, their alignment is said to be *interactive*.

Alignment in conversation was first studied in simple communicative situations in which participants were asked to describe spatial locations within a maze to each other (Garrod & Anderson, 1987). The main finding in these studies is that speakers tend to converge on the type of descriptions used to guide each other's movements through the maze. However, this convergence is not the result of speakers "grounding" their descriptions in the conversation by negotiating the terms they use (Clark & Wilkes-Gibbs, 1986). Rather, according to the alignment account, speakers start by simply providing descriptions of the spatial layout of the maze to each other. These descriptions are used to prime each other's mental models of the maze, which only then results in both speakers using similar types of description. (Garrod & Anderson, 1987). Because of this priming of mental models, mutual understanding is often achieved rapidly and without explicit negotiation between speakers.

Besides the sort of semantic or lexical alignment found in these early studies, other types of alignment have been reported at many other linguistic levels. For example, speakers describing pictures to each other reuse previously heard constructions, suggesting that speakers align their syntactic constructions (Branigan et al., 2000 e.g., "the girl gave the boy the book", instead of "the girl gave the book *to* the boy"). Furthermore, this has also been shown to be the case for phonetic features, like speech rates (Schultz et al., 2016; Ostrand & Chodroff, 2021). Lastly, alignment can occur in other modalities, as speakers tend also to align their gestures when talking to each other (Holler & Wilkin, 2011).

To get a better understanding of how alignment happens in dialogue, one needs to spell out two key assumptions that are at the core of this account. These assumptions are important because they also tell us in what sense this account sees communication as a joint action (Garrod & Pickering, 2009). The first assumption is that communicative acts, like speaking, signing, pointing and so on, belong, ultimately, to a larger category: that of actions. As a consequence, the cognitive mechanisms that allow us to perform actions, like reaching for objects, can also be used in the production of communicative acts, like pointing (Rizzolatti & Arbib, 1998).

The second assumption, proposed in earlier versions of the alignment account (Garrod & Pickering, 2004), involves a "parity" of the mental representations used in the production of one's own actions and in the understanding of someone else's actions. Put simply, when we observe others perform an action, like a reach towards an object, we recruit the same mental representations that are used when we prepare and execute a similar action ourselves (Hommel et al., 2001; Prinz, 1997). In terms of underlying mechanisms, this is often argued to involve an internal simulation of the other person's actions (Wilson & Knoblich, 2005; I. M. Thornton & Knoblich, 2006; Wolpert & Flanagan, 2001). In cognitive neuroscience, this phenomenon has been ascribed to the involvement of the mirror (neuron) system (Rizzolatti & Sinigaglia, 2010), whose activation during the performance and observation of actions has been found in monkeys (di Pellegrino et al., 1992) as well as in humans (Mukamel et al., 2010).

With these two assumptions in mind, we can understand how interlocutors come to align their communicative acts during dialogue. Since communicative acts are actions and, as such, both their production and comprehension rely on the same mental representations, then two people having a conversation (and thus producing and comprehending communicative acts) will come to activate the same mental representations and will, as a consequence, end up aligning their mental representations of the situation (Pickering & Garrod, 2013).

Alignment, from this perspective, provides another way of understanding communication as a joint action, where the focus is not so much on the principles that people use to achieve mutual understanding, but on the underlying cognitive mechanisms that make communication possible. Since these are the same mechanisms that enable people to perform and understand actions (including, importantly, nonverbal joint actions) the account represents, next to Clark's grounding account, one of the most explicit attempts at defining communication, and particularly spoken dialogue, as a joint action in and of itself.

More comprehensive versions of the interactive alignment account have complemented the mechanism of internal simulations of someone else's actions with a second mechanism: prediction (Garrod & Pickering, 2009; Pickering & Garrod, 2013). This more recent version argues that the internal simulations we run when observing other people's actions are, ultimately, at the service of predicting their most likely outcome; in other words, their intended goals (see forward models, Wolpert & Kawato, 1998; Wilson & Knoblich, 2005). Accordingly, communication is made possible not only because our representations are aligned in the sense of being similarly activated both during production and comprehension, but more importantly because these representations are being used to make real-time predictions about other people's communicative acts (Pickering & Gambi, 2018).

There are many communicative situations in which people rely on predictions of each other's communicative acts, but the one that has been extensively analyzed is when speakers take turns in conversation (Levinson, 2016). It has been noted that human conversations are structured around an ordered sequence of actions (i.e. turns), distributed across two or more speakers, with extremely short time gaps between each turn of approximately 200 ms (Sacks et al., 1974; Stivers et al., 2009). Being able to achieve such short gaps is an impressive feat, especially if we consider the fact that the time it takes to produce a single word is much longer than this, falling anywhere between 600 and 1000 ms (Levinson & Torreira, 2015). Thus, for hearers to respond rapidly, they need to prepare an appropriate response in advance, before the end of the speaker's turn (De Ruiter et al., 2006). In other words, they need to predict both what the speaker is saying (i.e, the speech act), as well as the likely ending time of what the speaker is saying, all the while starting to formulate the appropriate response (Magyari & de Ruiter, 2012). Several studies have shown that speakers continuously predict each other's turn durations by relying on a number of cues: linguistic (e.g., semantic/syntactic: Riest et al., 2015; prosodic: Bögels & Torreira, 2015; gestural: Kendrick et al., 2023; and facial: Nota et al., 2021).

In sum, by looking at the mechanisms that underlie dialogue, the interactive alignment approach has provided another way of looking at communication as a form of joint action, where the mechanisms that underlie joint action in general (i.e., simulation and prediction) are used during communication. Moreover, this approach has led scholars in the field to acknowledge the different ways in which low-level automatic processes, like priming and prediction, interact with higher-level, more intentional ones, like grounding (Knoblich et al., 2011; Rasenberg et al., 2020). One interesting way in which these two levels interact is found in the different strategies that people use to solve coordination problems when engaging in joint actions. I will now turn to one such strategy, and thus shift the focus from looking at communication as a form of joint action, to communication as a tool to facilitate the coordination of joint actions and the achievement of joint goals.

1.2 Communication as "coordination smoother"

When people perform a joint action, they often need to agree on the particulars of their individual contributions and on their coordination. For instance, when two people assemble a TV stand, they need to decide on who will hold the shelf and who will take care of fixing it to the wall, while also agreeing on the time and the place in which their actions will take place

(Clark, 2005). One way people usually solve these kinds of coordination problems is by relying on "coordination smoothers" that have the function to facilitate coordination (Vesper et al., 2010). Co-actors use different coordination smoothers during joint actions, from making their behavior more predictable (Vesper et al., 2011) to delimiting their action spaces to avoid bumping into each other (Vesper et al., 2009). A further example of a coordination smoother is, of course, communication (Clark, 1996).

When seen as a coordination smoother, communication becomes a means to an end, where the end in question is to facilitate the achievement of other joint goals, outside of the specifically communicative goal of understanding each other (Clark, 2005; Vesper et al., 2010). Below, I review a primary example of communication used for these purposes, which will occupy us for the rest of the Introduction and will be the focus of the present thesis.

1.2.1 Sensorimotor communication

Although people often resort to language to facilitate coordination, many joint actions require such high levels of temporal and/or spatial precision in their execution that spoken language is simply inadequate. For example, when two pianists need to synchronize with precision the onset of their playing, talking to each other is either not possible (e.g., during a live music performance), or too inefficient (e.g., because it takes too long to describe certain actions with words) (Goebl & Palmer, 2009). A further complication comes from the fact that sometimes people performing joint actions simply do not speak the same language (Clark, 2005). When confronted with these situations (i.e., coordination problems), people can take advantage of the fact that the movements they perform as part of the joint action, like reaching for and passing an object, are being closely monitored by their co-actor (i.e., the intended recipient of the object). A number of studies on action observation suggest that people can derive a wide range of useful information from simply observing someone else's movements (for a review, see Becchio et al., 2012), including information that is relevant for attaining a joint goal (Manera et al., 2013). However, in many joint actions people perform these same movements but in such a way that it makes it easier for a co-actor to perceive and/or predict their intended goals. Concretely, people can exaggerate certain kinematic properties of their movements, such as their velocity, duration, or direction, and thereby inform co-actors about their intended goal. This kinematic information can then be used by co-actors to predict the action goal and adapt their own movements accordingly, often leading to better coordination.

The basic principle in this form of communication, first alluded to by Clark (1996, 2005) and then further developed by Pezzulo and colleagues (2013, 2019), is that instrumental actions, when they systematically deviate from their most efficient performance, become more predictable and discriminable to co-actors in a joint action, as well as to non-interacting observers (Vesper & Sevdalis, 2020). In other words, by deviating from efficient performance, actors can add a communicative function on top of the instrumental function of their action (Dockendorff et al., 2019; Sebanz & Knoblich, 2021). From this perspective, "sensorimotor communication" (henceforth SMC; also known as "action-based communication", Sebanz & Knoblich, 2021) differs from other types of communication, particularly verbal communication, in the following way: the information 'channel' used for communication is the same as the one used to perform the instrumental action (Pezzulo et al., 2013, 2019). Thus, it is *while* one of the pianists exaggeratedly raises his arms above the left side of the keyboard that he *both* prepares to strike one of the lower keys (thereby stating his part of the duet, i.e., his instrumental goal), *and* informs the other pianist about the exact moment he will do this (i.e., his communicative goal) (Goebl & Palmer, 2009). Following previous terminology, I will use the term "communicative modulations" (of instrumental actions) to refer to these actions (Schmitz et al., 2018b; Vesper, Schmitz, et al., 2017; Trujillo et al., 2018).

The fact that people can produce communicative modulations, and thereby combine instrumental and communicative goals within a single action, makes this form of communication extremely useful in joint actions. This is because in many joint actions co-actors need to exchange information *while* they are acting together, but cannot rely on (or simply do not have) more conventional means to do this, from playing in an orchestra (during a live concert) to walking together (while simultaneously having a conversation). Because of this, SMC is a prime example of communication being used as a means to facilitate joint actions (i.e., it serves as a coordination smoother) (Vesper et al., 2010).

Apart from its role in facilitating coordination, SMC is an interesting form of communication because it relies on signals whose "meaning" is readily understood by co-actors, unlike in forms of communication that are based on conventions. As I will argue more extensively below, the ease of understanding meaning in SMC is based on people's tacit knowledge of motor acts that they themselves perform or that they see others performing on a daily basis, such as reaching, pressing, grasping, placing, giving and so on. Thus, when modulating their actions communicatively, co-actors draw on this tacit knowledge about actions and they can safely assume that others will draw on it as well. This then provides enough common ground for co-actors to start exchanging information rapidly, without the need for extensive negotiation about the meaning of the signals used (Pezzulo et al., 2019; Schmitz et al., 2021).

Like any other form of communication, SMC requires, for it to be successful, receivers who can detect the signals sent by the sender. In the case of SMC, observers need to be able to detect kinematic deviations present in their co-actors' movements. Can observers do this? Studies on action observation, where participants are presented with videos of non-exaggerated instrumental actions and are asked to predict the goal of the actor, suggest that they typically can do so (see Becchio et al., 2012 for a review). This includes the capacity to distinguish between action goals like grasping-to-pour or grasping-to-drink (Cavallo et al., 2016), social goals like cooperating or competing with another person (Manera, Becchio, Cavallo, et al., 2011), and even the goals of two people either interacting with each other or acting in parallel (Manera et al., 2013).

The capacity to predict goals based on movement kinematic, likely relies on internal simulations of the observed action, from which the observer can generate predictions about the agent's most likely goal (Wilson & Knoblich, 2005; Wolpert & Flanagan, 2001). But what about actions that deviate from their most efficient performance, as in the case of SMC? In many cases, communicative modulations of instrumental actions have the effect of facilitating the predictions made by these internal simulations. This then makes it easier for a co-actor to predict different aspects of the movement, such as the spatial goal as well as the temporal dimension (Pezzulo et al., 2013; Wolpert et al., 2003).

1.2.2 SMC - Previous research

Given its key role in facilitating coordination, SMC has mostly been studied in joint tasks in which one participant informs a co-actor about an aspect of the task that either enables or facilitates the achievement of a joint goal (but see Vesper, Schmitz, Safra, et al., 2016 for an example where both co-actors use SMC synchronously). For example, pairs of participants are asked to coordinate their actions to simultaneously grasp an object (Sacheli et al., 2013), or aim for the same target location (Pezzulo & Dindo, 2011). However, the information needed to achieve the joint goal (e.g., information about the part of the object that needs to be grasped, or the target to be aimed at) is available to one member of the dyad only, the "Leader" participant. Furthermore, Leaders are usually not allowed to talk to "Follower" participants, thus posing an obvious coordination problem for the dyad, namely, how can Leaders, without talking, inform Followers and thereby successfully attain the joint goal?

Previous studies show that, to solve this problem, Leaders spontaneously exaggerate the kinematics of their instrumental actions to convey the required information to Followers. To illustrate, one study asked pairs of participants to perform a sequence of tapping movements directed at different targets while instructing them to coordinate each tap both spatially (i.e. tapping on the same target) and temporally (i.e., tapping the target at the same time) (Vesper & Richardson, 2014). At the beginning of each trial, the Leader received information about the correct target, whereas the Follower did not receive any information but was able to observe the Leader's arm movements. The results showed that Leaders systematically modulated the amplitude of their arm movements, which made it easier for Followers to discriminate between target locations, leading to higher coordination success (for a review, see Pezzulo et al., 2019).

There is one important aspect about previous studies on SMC that I would like to highlight. This relates specifically to how Leaders establish a communication system while relying on "iconic" mappings between movements modulations and goals. These mappings have been described as iconic because they are governed by regularities found in the way people normally perform certain movements (Fitts & Peterson, 1964; Jeannerod, 1984). For example, in Sacheli et al., (2013) Leaders established a communication system using both the size of their grip (i.e., precision or power grip) and the amplitude of their reaching movement (i.e., higher or lower) to inform Followers about particular locations of a bottle-shaped object they both needed to grasp synchronously. Grasping the *upper* part of the object required a *precision* grip directed at a *higher* location and, conversely, grasping the *lower* part required a *power* grip directed at a *lower* location. By modulating these two movement parameters (grip size and amplitude) Leaders relied on the regularities present in their movements when the normally execute reaching actions. In short, they produced "iconic" mappings.

Similarly, in Vesper et al. (2017), Leaders conveyed information about target locations by creating stable iconic mappings between the duration of their aiming movements and the relative distance of the targets. Specifically, they increased the duration of their movements if the target location was further away, while they decreased the duration when the target location was closer. Since aiming movements normally take longer when they are directed at far locations, relative to near ones (Fitts & Peterson, 1964), these mappings are, again, “iconic”.

1.3 Research questions

My main goal in this thesis is to expand the focus of SMC by asking whether people can rely on communicative modulations that convey information about action goals that are both spatially and temporally separated from the action leading to their achievement. I will call these goals “distal goals”. I will explain what I mean by distal goals and how they contrast with more “proximal” ones shortly.

To address this question, I will present three empirical studies focusing on the *observation* (**Chapters 2** and **3**) and the *production* (**Chapter 4**) of communicative modulations. In what follows, I will present the questions that each of these empirical studies is aimed at addressing and spell out the theoretical motivation underlying them.

1.3.1 From proximal to distal goals in sensorimotor communication

In **Chapter 2**, I start by arguing that most research in SMC has focused on communicative modulations that convey information about “proximal goals”. Proximal goals are the goals that are immediately achieved by an action, and are thereby directly tied, both temporally and spatially, to this action. For convenience, I call the actions leading to proximal goals “proximal actions”. In the specific case of SMC, these proximal goals include goals such as grasping an object (Sacheli et al., 2013; Candidi et al., 2015) or aiming at a target location (Vesper & Richardson, 2014; Vesper, Schmitz, et al., 2017).

In some occasions, proximal goals are subgoals of more future, or as I will call them in this thesis, “distal” goals (Marteniuk et al., 1987). A reaching action, for example, is sometimes performed with the further distal goal of manipulating, moving, or showing the object to someone. As such, distal goals are often the result of executing two or more proximal actions, one after another, leading thus to an *action sequence* (Marteniuk et al., 1987).

In Chapter 2 I review a number of studies in motor control and in action observation showing that distal goals can have reliable and visible effects on the kinematics of early proximal actions. All this research, however, is limited to instrumental (non-communicative) actions, and thus leaves open the question of whether observers can interpret communicative modulations of early proximal actions in terms of their most likely distal goals. This is the main question I ask in Chapter 2.

In asking whether observers can predict distal goals, I introduce the term “motor-iconicity” to describe one of the mappings that participants produce to connect early communicative modulations to distal goals. Similar to “iconicity” in the case of proximal goals (see 1.2.2 above), I use “motor-iconicity” in this thesis to refer to a mapping that preserves kinematic differences found whenever people perform actions naturally. After introducing this notion in Chapter 2, I ask whether observers have a preference for motor-iconicity when mapping communicative modulations of an early action to a distal goal.

1.3.2 Simulations underlying the prediction of distal goals

While Chapter 2 raised the question of *whether* observers can predict distal goals from early communicative modulations (and whether they do so motor-iconically), in **Chapter 3** I turn to the question of *how* they do this. Specifically, I ask under which conditions observers can predict distal goals when observing communicative modulations. To this aim, I focus on early actions that are part of a two-step action sequence, like the initial reaching component of a reach-to-give action sequence. When observers look at such an action and try to predict its distal goal, do they rely exclusively on kinematic information given away by the first (reaching) component? Or do they also rely on kinematic information of the second (giving) component? And are there situations in which one piece of kinematic information matters more than the other?

In Chapter 3 I try to provide an answer to these questions by proposing that observers can rely on two types of simulation when observing the first action step of a two-step action sequence. The first type of simulation, "movement-to-goal simulation", uses the kinematic information present in the first action step to directly predict the distal goal. The second, "movement-to-movement simulation", uses the kinematic information of the first action step to simulate the second action step and then, on the basis of this latter simulation, predict the distal goal. Thus, in Chapter 3 I ask whether observers rely exclusively on direct movement-to-goal simulations to predict a distal goal, or whether they sometimes also rely on movement-to-movement simulations.

1.3.3 Modulating one's early actions to communicate about distal goals

Chapters 2 and 3 ask about the observer side of communication, i.e., the "receiver". But what about the other side of it, i.e., the "sender" side? Are senders able to use communicative modulations to inform others about distal goals? And how do they do it (e.g., by relying on motor-iconic mappings)? In **Chapter 4** I turn to these questions. Specifically, I ask whether participants in the role of senders can modulate their actions in a joint action context to provide information about a distal goal to a co-actor. I do this by introducing a new joint task in which pairs of participants have to cooperate in performing a two-step action sequence directed at one of two distal goals. They are only able to see each other's movements on a screen and, importantly, only one of them knows which of the two distal goals is the correct one. This provides an ideal setting for senders to modulate their movements to communicate about a distal goal.

While the main focus of Chapter 4 is on the sender, the joint nature of the task allows me to further ask questions about their interaction more generally. Are receivers able to understand the communicative modulations? And do these modulations lead, in turn, to better coordination? Are senders and receivers able to establish a communication system based on movement modulations? And finally, how does this system develop as a result of their interaction?

In the last part of the thesis, I provide a summary of the main findings of the three empirical chapters and discuss a few theoretical implications for the study of SMC and joint action more generally. I conclude with a few take-away messages.

Chapter 2. From proximal to distal goals in sensorimotor communication

2.1 Introduction

People engage in a variety of complex social interactions that require temporal and spatial coordination of their individual actions, ranging from carrying a sofa with someone to performing a musical duet (Sebanz et al., 2006; Sebanz & Knoblich, 2021). In order to achieve such a feat, interaction partners often predict each other's actions by relying on behavioral cues from which they can derive useful anticipatory information about what their partners are about to do (Sebanz & Knoblich, 2009). These cues, in turn, facilitate interpersonal coordination and the achievement of joint goals. One type of cue, which has received particular attention during recent years, consists of communicative modulations of instrumental movements. For instance, two pianists playing a duet might lift their fingers higher, and by doing so, inform each other about the exact timing of their actions (Goebl & Palmer, 2009). Similarly, when carrying a sofa together, the person who is walking forward might exaggerate an upward movement with the sofa to inform the person who is walking backward about the upcoming staircase (Vesper, Abramova, et al., 2017). By exaggerating their movements in this manner, interaction partners can fulfill two goals simultaneously: an instrumental goal, such as playing a piano piece or moving a sofa, and a communicative goal of informing an interaction partner about one's goals and intentions. This general capacity to provide anticipatory information about one's goals and intentions by means of communicative modulations of instrumental movements has been termed *sensorimotor communication* (SMC) (Pezzulo et al., 2013).

A growing body of research has investigated SMC in experimental tasks in which two participants coordinate their actions to achieve a joint goal while the information relevant for attaining this goal was allocated asymmetrically between them (for a review, see Pezzulo et al., 2019). “Leader” participants with full task information have been shown to spontaneously modulate certain kinematic features of their goal-directed movements, such as grip aperture (Candidi et al., 2015), movement direction (Pezzulo & Dindo, 2011), movement amplitude (McEllin, Knoblich, et al., 2018; Vesper & Richardson, 2014), and velocity (Sacheli et al., 2013; Vesper, Schmitz, et al., 2017), to make their actions more informative, and hence more predictable for “Followers” participants who only have incomplete task information (Vesper, Schmitz, et al., 2017).

In order for SMC to be an effective form of communication, Followers need to be able to perceive the kinematic modulations in Leaders’ goal-directed movements. Growing evidence indicates that observers can indeed perceive such modulations (e.g., Becchio et al., 2008, 2012; Cavallo et al., 2016; Manera, Becchio, Cavallo, et al., 2011; McEllin, Sebanz, et al., 2018), and that they can understand them as conveying specific information about the Leader’s proximal (i.e. immediate) goals, such as reaching for a particular object (Pezzulo & Dindo, 2011) or aiming towards one of several target locations (Vesper & Richardson, 2014).

What allows for such sensitivity towards others’ actions is the fact that observers use their own motor systems to predict others’ unfolding actions and goals by generating internal simulations of the observed movement (e.g., forward models, see (Desmurget & Grafton, 2000; Wilson & Knoblich, 2005; Wolpert et al., 2003). The predictive nature of such internal simulations enables the observer to revise and update her expectations about the actor’s goals in a timely fashion, particularly in cases where the observed movements deviate from their most efficient performance, as is the case in SMC (Pezzulo et al., 2013; Trujillo et al., 2018). These latter cases where actors deviate from efficient performance have been argued to constitute a proper form of communication to the extent that they enable observers to both derive useful anticipatory information about an actor’s goals and to interpret and disambiguate between different, sometimes competing, goals (Pezzulo et al., 2019). Accordingly, we regard actions as communicative when they have the potential to facilitate an observer's prediction of another actor's upcoming goals. Note that by adopting this broad definition of communication, we focus on the receiver end of the interaction, i.e. how the receiver reacts to such communicative actions. In contrast, other perspectives on communication tend to highlight the role that the sender plays in producing communicative actions and they often require mutual awareness of communicative intentions in both producing and understanding these actions (Sperber & Wilson, 1995). The broad definition on which we base our work is consistent with previous findings in SMC showing that, by interpreting a particular kinematic modulation as conveying specific information about an actor's goal, observers can adapt their own behavior in ways that facilitate interpersonal coordination and the achievement of a joint goal (Candidi et al., 2015; Sacheli et al., 2013; Vesper, Schmitz, et al., 2017).

What is less clear from previous research is whether the use of SMC is restricted to facilitating predictions of immediate proximal goals or whether individuals can also interpret movement modulations that encode information about their partner's upcoming *distal goals*, i.e., goals that go beyond the observed action and thus are only attained after the achievement of a more proximal (sub-) goal first. The aim of the present study was to extend previous research on SMC by focusing on how observers interpret communicative modulations of instrumental actions that convey information not only about proximal, but also about distal goals.

2.1.1 From proximal to distal goals

To illustrate the difference between proximal and distal goals, consider a situation where a football player (Player A) recovers the ball on her side of the field and prepares a quick counterattack. Two of her teammates (Players B and C) start running along the flanks towards the opposite goal, ready to receive the ball. At this point, Player A could simply pass the ball to either of her teammates, thereby fulfilling her proximal goal. An alternative would be for Player A to continue dribbling the ball up to the midline, and to only then pass it on to one of her teammates. In this latter situation, the dribbling of the ball has become the more proximal goal, while the passing of the ball is now the more distal goal, since it follows temporally and is mediated by the prior achievement of a more proximal goal.

Although proximal and distal goals are separated in time, there is now strong evidence showing that distal goals can affect the kinematics of early components of proximal actions. For example, when individuals perform reach-to-grasp movements towards an object, different distal goals (e.g., throwing the object into a large box or placing it in a well) differentially affect the velocity of the early transport phase of the movement (Marteniuk et al., 1987). Relatedly, when participants perform two-step action sequences, the specific constraints imposed on the second action component (e.g., pouring from a bottle or throwing it) can influence the kinematics of the first component (e.g., grasping the bottle) (Cavallo et al., 2016; Lewkowicz & Delevoye-Turrell, 2020; Rand et al., 1997; D. Rosenbaum et al., 1990). These findings can be interpreted in terms of a more general binding procedure that links both motor and perceptual features of a distal goal when organizing multiple movement segments within a "common event file" (Hommel et al., 2001). As a consequence of this binding, the activation of relevant perceptual features of a distal goal can lead to the concurrent activation of the appropriate motor program that is normally used to achieve that goal (Hommel, 2009; also see Fogassi et al., 2005 for a similar argument, but supported by neurological evidence).

Similar "backpropagation effects" have been reported in social tasks, where distal social goals (e.g., passing an object to another person or placing it in front of her) have been shown to affect the kinematics of early action components (e.g., reaching towards the object, Ansuini et al., 2008; Becchio et al., 2008; Georgiou et al., 2007). With respect to the perception and interpretation of these movements, recent findings suggest that observers can extract and use these early kinematic cues to discriminate between actions performed with different social intentions (Manera, Becchio, Cavallo, et al., 2011; Sartori et al., 2011) or to predict the outcome of a ballistic movement, such as when someone is throwing a ball (Maselli et al., 2017).

The question that these studies leave open is whether such early effects of distal goals on the kinematics of proximal movements could also be used communicatively, i.e., in cases where agents intentionally make their distal goals easier to predict. To illustrate this idea, consider again our football example from above and assume that Player A wishes to inform her teammates that she will pass the ball to Player C, who is already much further down the field than Player B. To make this intention explicit to her teammates, Player A visibly increases her movement speed, thereby demonstrating that she is preparing for a long, powerful pass to Player C (rather than a short pass to Player B). By increasing her dribbling speed, Player A modifies the kinematics of an early action aimed at achieving a proximal goal in a way that provides useful information about her future distal goal to her two teammates. Importantly, the two teammates can use these early kinematic cues to predict and disambiguate between possible distal goals, and adapt their behavior accordingly, e.g., Player C can prepare to receive the ball, whereas Player B can try to draw the attention of the opposite's team defenders away from the passing sequence.

A recent computational account of early intention recognition of sequential actions formalizes the idea that observers can disambiguate between an observed agent's distal goals early, but only when the agent *co-articulates* the two movement primitives within the sequence (Donnarumma et al., 2017). Co-articulation, in this context, means that the agent alters the execution of an earlier proximal movement (e.g., reaching and grasping a bottle) in order to satisfy the specific constraints posed by the achievement of an upcoming, more distal goal (e.g., pouring from the bottle or simply moving it). Through a series of computational simulations as a proof-of-concept, Donnarumma and colleagues showed that when two sequential proximal movements are co-articulated, the kinematic features of the first movement are sufficient for an observer agent to correctly identify and disambiguate the distal goal. Importantly, their proposal also put forward the possibility that co-articulation might be used by actors strategically, as a way of helping an observer understand their distal goals (see the Appendix of Donnarumma et al., 2017). Here, we address this possibility empirically by drawing on a) previous research showing that distal goals can affect early action components (e.g., Lewkowicz & Delevoye-Turrell, 2020) and b) computational simulations suggesting the possibility to use early kinematic cues to disambiguate between distal goals (Donnarumma et al., 2017).

2.1.2 Motor iconicity in SMC

Although the research reviewed above suggests that observers might be able to use early action components to predict an upcoming distal goal, it leaves open the question of how communicative modulations of those same actions might be interpreted with respect to different distal goals. As highlighted earlier, movement deviations not only make proximal goals easier to predict, but also allow observers to disambiguate between potential action alternatives. For example, previous studies have shown that specific kinematic features of a Leader's movement, such as movement duration, height or direction, can be used to disambiguate between target locations that differ in terms of distance (near or far; Vesper, Schmitz, et al., 2017), height (upper or lower; Sacheli et al., 2013) or location (left or right; Pezzulo & Dindo, 2011). Such systematic mappings between kinematic features of movements and proximal goals have recently been described as "iconic" (Vesper, Schmitz, et al., 2017), because the relation between them involves some form of similarity which can be easily identified by both senders and receivers (Allwood, 2002). In the case of SMC, the relevant similarity is established between a particular movement used by the sender and the *most likely* goal that such movement achieves during natural performance. For example, a *higher* movement trajectory communicates a *higher* final grasp location (Sacheli et al., 2013) - because higher grasping movements are naturally performed with higher movement amplitude (also see (Schmitz et al., 2018b). A similar observation regarding the similarity between communicative movements and their goals has been made by researchers in sign language, who have argued that verbs in American Sign Language (ASL) used to refer to the manipulation of a tool (i.e., so-called handling classifier verbs such as BRUSH-HAIR or BOUNCE-BALL) are represented "motor-iconically" with a handshape that depicts how a person grasps and manipulates the tool, as well as the movements typically performed with it (e.g., brushing one's hair or bouncing a ball) (Emmorey et al., 2004). Similarly to the case of SMC, the particular handshape and movements used in ASL to represent a tool correspond to the hand movements and goals that one would normally achieve while manipulating it. In line with these observations, we will henceforth refer to the relation between movements and their most likely goals as "motor-iconic".

In the case of distal goals, the motor-iconic relation between movements and goals can be captured by looking at the specific ways in which people perform an action when this same action is directed at a proximal goal. For example, when people perform unconstrained aiming movements towards far proximal targets they use higher peak velocity compared to near proximal targets (Jeannerod, 1984). If this relation is invoked in someone who observes such movements in order to predict an agent's distal goal, then it can equally be considered as motor-iconic. In other words, the motor-iconic relation underlying the observation of movements and their distal goals is grounded on an understanding of that same movement were it to be directed towards a proximal goal.[1]

[1] Note that this way of defining iconicity departs in some ways from more standard definitions in semiotics and linguistics that focus on the relation between the form of a sign and a referent in the world (Wilbur, 1987). Instead, motor-iconicity focuses on the relation between a particular form (e.g., a movement or a gesture) and the contents of a motor representation in the mind of the speaker or signer (e.g., a goal) (Emmorey, 2014; Taub, 2001).

Taken together, these considerations lead to the specific empirical question of whether observers map proximal communicative modulations of goal-directed movements onto proximal and distal goals in a motor-iconic manner, i.e., in a manner that preserves the link between the kinematic features of the proximal movement and its most likely proximal or distal goal.

2.1.3 The present study

In a computer-based online experiment, participants observed animations of a box being moved at different velocities along a horizontal line from a start location towards a designated movement endpoint. Due to a partial occlusion of the visual scene, participants were not able to observe how the box actually reached the target location. They could therefore only rely on features of the observable proximal part of the movement to determine the likely final location of the box. After observing the animated movement, participants were asked to select the target location which they considered the likely proximal or distal goal of the action (Figure 1). In contrast to previous studies, in which participants observed movements in two- or three-dimensional space, participants in the present study observed animations of one-dimensional sliding movements in order to single out the role of temporal movement parameters (e.g. velocity and duration) for extracting information about the target locations.

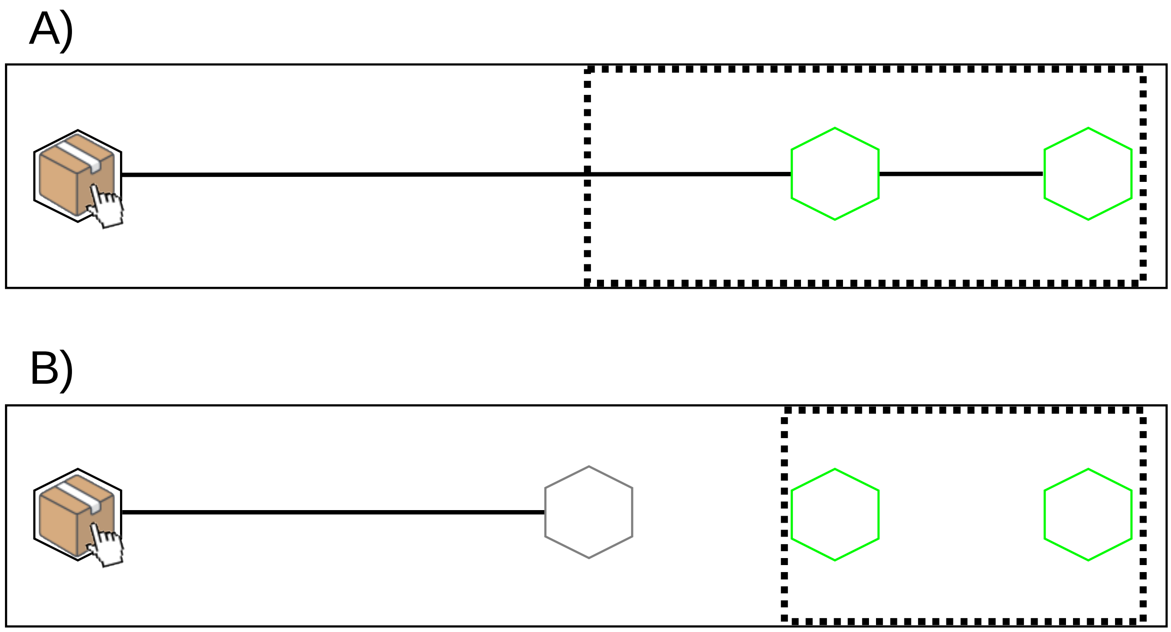

Figure 1. Experimental layout used in Experiments 1-3.
Experimental layout in the (**A**) Proximal goal condition and the (**B**) Distal goal condition. The black dotted line represents the outline of the occluded area during trials, where the (near and far) target locations are displayed in light green.

To find out whether observers extract information about distal goals from early kinematic modulations, we manipulated whether the animated movements that participants saw achieved a proximal goal only (i.e., sliding a box to one of the two target locations) (Figure 1A: "Proximal goal" condition) or achieved a distal goal (i.e., delivering the box to one of the target locations) by means of achieving a proximal goal first (i.e., sliding the box towards a middle target) (Figure 1B: "Distal goal" condition). In this latter condition, the achievement of the distal goal was made possible by having the box disappear from the display and reappear within one of the target locations. The purpose of this "teleportation" was to introduce a visible spatial and temporal separation between the proximal movement and the distal goal. Such separation prevented participants from simply extrapolating the proximal sliding movement towards the distal goal since the movement endpoint was now shifted away from the target locations towards the middle of the display. Finally, this separation also enabled us to keep the

Proximal and Distal goal conditions as similar as possible as both contained only a single sliding movement that participants needed to interpret.

Drawing on previous research on SMC, we had three central predictions: First, we predicted that participants would be able to detect differences in the velocity of the observed movements, and that this detection would allow them to consistently map these movements to one of the two potential goals (i.e., target locations). Second, we predicted that the stronger the communicative modulation of velocity, the easier participants' decision should be, resulting in higher consistency of their mappings. Third, based on the lawful relation between movement velocity and distance of natural movements, where farther target locations are reached with higher peak velocities in unconstrained aiming movements (Jeannerod, 1984), we expected participants to map faster movements onto the far target location and slower movements onto the near target location. This mapping would represent what we call a motor-iconic relation, as it preserves the underlying link between observed movements and their most likely goals.

2.2 Experiment 1. Interpreting velocity modulations

The aim of the first experiment was to establish the experimental paradigm by testing our main predictions regarding the role of distal goals in interpreting modulations in the velocity of proximal movements.

2.2.1 Methods

Participants. We recruited 50 participants (16 women; Age: M = 29.6 years; SD = 9.9 years), 25 per condition, through the online testing platform Testable (https://www.testable.org/). Sample size was determined using the Superpower statistical package (Lakens & Caldwell, 2021) on R Studio (R Core Team, 2013). We aimed at obtaining a medium effect size (.4) and high statistical power (>.8) based on a series of well-established findings showing that participants can detect subtle kinematic cues to predict other agents' goals (Becchio et al., 2008; Cavallo et al., 2016).

Participants were all proficient English speakers, and were paid 1.5£ for an estimated study completion time of 10 minutes. All participants gave prior written informed consent in accordance with the United Ethical Review Committee for Research in Psychology (EPKEB). This design and analysis of this study was pre-registered at https://osf.io/2qkn3. All data and materials are available on OSF, at https://osf.io/pv74b.

Stimuli. The basic layout for each experimental condition of Experiment 1 is shown in Figure 1. In both conditions, participants saw a stationary box with a mouse cursor attached to it. The box and cursor were displayed within a black hexagonal location on the left side of the screen. During familiarization, participants also saw two green hexagonal target locations on the right side of the screen. During trials, these two target locations were covered with a rectangular black occluder.

A black horizontal line, along which the box moved during the trials, connected the initial location to the green target locations on the right side of the screen (Proximal goal, Figure 1A), or to a grey hexagonal area in the middle of the screen (Distal goal, Figure 1B).

The animations of the box movements were created from averaged actual mouse movements recorded with PsychoPy (J. W. Peirce, 2007) in a setup identical to the one shown in Figure 1A (Proximal goal condition). Movements to near and far targets were averaged separately, thus obtaining two "prototypical" natural movements, one for each target location (henceforth "Normal near" and "Normal far" movements). Based on the two Normal movements, exaggerated movements were generated by, first, identifying the peak velocities for each (i.e., Normal near and Normal far) movement. Then, we rescaled both of these movements such that the peak velocity was either one or two standard deviations below the peak velocity of the Normal near movement, or one or two standard deviations above the peak velocity of the Normal far movement. This procedure led to overall six different movements (Figure 2A): two “Normal” ones (i.e., Normal near and Normal far), two “Exaggerated” ones (i.e., Slow and Fast) and two “Very exaggerated” ones (i.e., Very slow and Very fast).

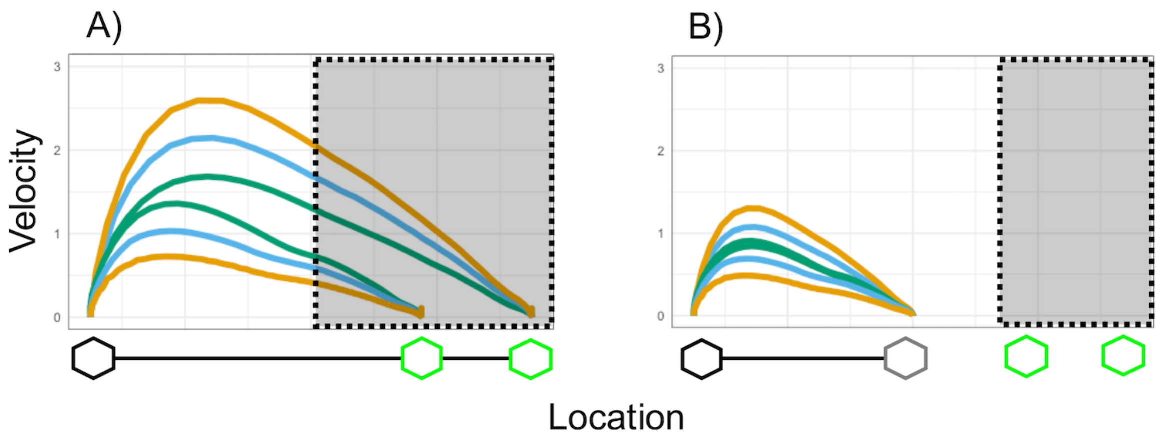

Figure 2. Velocity profiles of sliding movements

Velocity profiles of sliding movements used in the (**A**) Proximal goal condition and the (**B**) Distal goal condition. Normal movements are colored in green, Exaggerated movements in blue, and Very exaggerated ones in yellow. The dotted areas represent the occluded areas in each condition during experimental trials.

To create the movement animations in the Distal goal condition, these six movements were further reshaped so that their endpoints would all converge towards the middle of the screen (Figure 2B). Critically, this procedure retained most kinematic features of the original movements (e.g., bell-shaped velocity profiles) but eliminated the differences in movement distance such that all movements now had the same endpoint in the middle of the screen. Further details on how movements were recorded and averaged are provided as in the Supplementary Material, including details on how the rescaling affected the movements in the Distal goal condition.

Design. The experiment consisted of a mixed factorial design, with one between-subject variable (goal type) and one within-subject variable (degree of exaggeration). The between-subject variable manipulated whether the sliding movement achieved a proximal or distal goal. Specifically, the proximal goal consisted in simply reaching one of the two target locations (i.e., green targets) in the Proximal goal condition (Figure 1A and 2A). In the Distal goal condition, the endpoint of the sliding movement was temporally and spatially separated from the final target locations to an intermediate target located in the middle of the screen. This meant that the goal of reaching one of the two target locations was only achieved by moving to this intermediate target first (Figure 1B and 2B). As a consequence, reaching one of the two actual target locations became the distal goal in this condition, while moving towards the middle target became the proximal one. The within-subject variable manipulated whether and to which degree the animated movements were exaggerated in terms of peak velocity (i.e., Normal (i.e., no exaggeration), Exaggerated, Very exaggerated).

Procedure. After being randomly assigned to one of the goal type conditions, participants were familiarized with the task. They were first presented with the complete task layout (as illustrated in Figure 1), but without the occluder covering the green target locations. Participants in both conditions then saw two successive Normal movements of the box, one to the near target, the other to the far target (order counterbalanced across participants) (see Figure 2, Normal movements in green). In the Proximal goal condition, participants saw the box moving at a Normal velocity to one of the two green target locations (see Figure 2A). In the Distal goal condition, after participants had observed the box moving at Normal velocity to the middle of the screen, the box disappeared for approximately 500 ms and then reappeared in one of the two green target locations (note that in this condition the two Normal movements had overlapping velocity profiles, see Figure 2B and Supplementary Material for further details). After seeing these two movements, participants in both conditions were asked to select the target location where they saw the box had moved by pressing the “n” key (for near) or “f” key (for far).

Next, a black occluder covered the target locations and participants were told that during the actual experiment, they would be presented with another participant's previously recorded movements. Importantly, they were informed that this previous participant had produced the movements "in ways that would help others guess to which green target location he/she was moving the box". This information was provided in order to make it explicit to participants that the movements they were about to see were communicative, that is, that they contained useful information about the previous participant's goals.

Participants performed 36 experimental trials, divided into six blocks. In each trial, they were presented with an animation of the box sliding along the black line to either the occluded target locations in the Proximal goal condition, or towards the middle of the screen in the Distal goal condition. In this latter condition, once the box reached the middle of the screen, it remained stationary for a few moments, and then disappeared from the display. Participants in both conditions were then prompted to answer to which location they thought the box had been delivered. A trial was completed when participants pressed one of the two assigned keys ("n" or "f"), corresponding to either the "near" or "far" target locations. Each block contained all six degrees of exaggeration, presented in random order. Participants did not receive feedback about their performance at any point. At the end of the experiment, participants were asked to fill out a short questionnaire about their experience with the task.

2.2.2 Results

Data preparation. We categorized participants' responses as *Iconic* or *Non-Iconic* mappings. Iconic mappings refer to trials where participants pressed the "n" key in response to movements with lower peak velocity and the "f" key in response to movements with higher peak velocity. Note that Iconic mappings correspond to the "motor-iconic" relation, since they preserve the relation between movement velocity and most likely goal. Non-Iconic mappings, on the other hand, refer to those trials where participants reversed this association, i.e., by pressing “f” in response to movements with lower peak velocity and “n” in response to movements with higher peak velocity.

Two dependent variables were computed from participants’ number of Iconic and Non-Iconic mappings (aggregated across all six blocks): Calculating the *absolute* difference between the total number of Iconic mappings and the total number of Non-Iconic mappings, separately for each goal type condition and each degree of exaggeration, gave us a *Consistency score* for each participant ranging from 0 to 12. A Consistency score of 0 meant that participants mapped velocities randomly to targets and a score of 12 meant that participants mapped with absolute consistency. Calculating the *signed* difference between Iconic and Non-Iconic mappings gave us the *Mapping score*, which could range from +12 (fully iconic mappings) to -12 (fully non-iconic mappings). A Mapping score of 0 meant that participants lacked a preference for Iconic or Non-iconic mappings.

Participants who pressed the same key (either “n” or “f”) at least ten times in a row were excluded from further analysis. Based on this criterion, one participant was excluded in Experiment 1.

Mapping consistency. To test whether participants interpreted the observed velocity differences in a consistent manner, we compared the distribution of Consistency scores to 0 (i.e., inconsistent mapping) using separate Bonferroni-corrected one-sample *t*-tests. Consistency scores are displayed in Figure 3, where each dot represents an individual participant grouped according to degrees of exaggeration. The scores differed significantly from 0 across all degrees of exaggeration and goal type (all $t(23) > 6.4$, $p < .001$, $d > 1.3$, one-tailed). This result shows that participants were able to distinguish the different animated movements in terms of velocity and, thereby, to consistently map them to either the near or the far target location, regardless of how exaggerated the velocity profile was and regardless of whether the movement achieved a proximal goal or a distal goal.

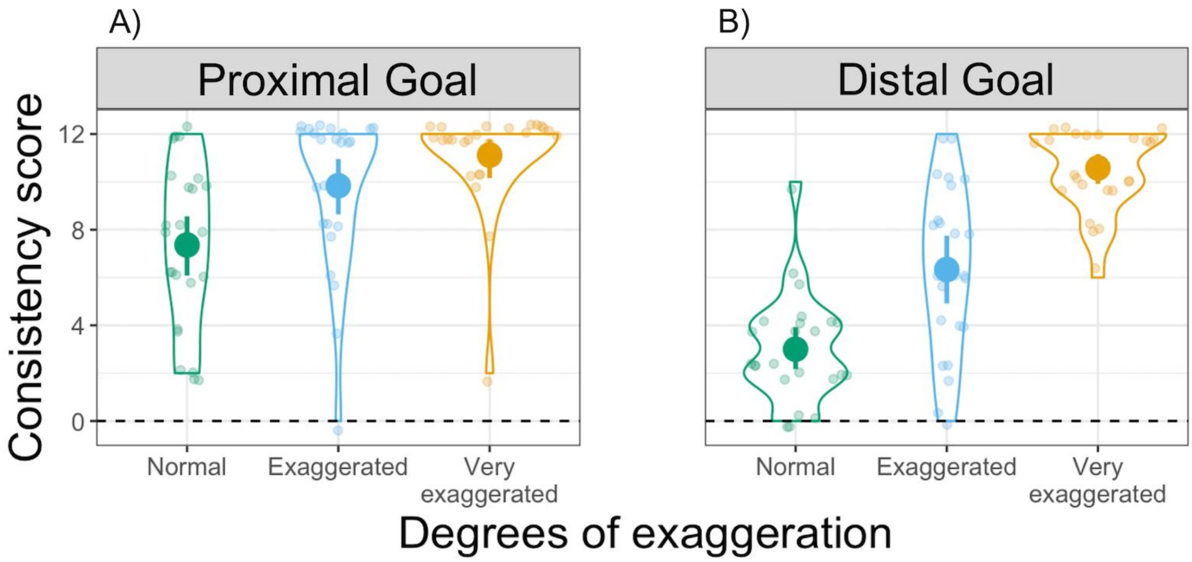

Figure 3. Consistency scores in Experiment 1.
Distribution of Consistency scores in the (**A**) Proximal goal and (**B**) Distal goal conditions. Each dot represents an individual participant, with one Consistency score for each degree of exaggeration: Normal in green, Exaggerated in light blue and Very exaggerated in yellow. Violin plots represent the overall distribution of Consistency scores for each degree of exaggeration. The dashed horizontal line indicates a hypothetical value for random mapping (i.e., no consistency).

To address the role of exaggeration and goal type, we conducted a 2x3 ANOVA with Consistency scores as dependent variable, goal type (Proximal and Distal goal) as between-subject variable and degrees of exaggeration (Normal, Exaggerated, Very exaggerated) as within-subject variable. We found a significant main effect of goal type ($F(1,47) = 23.6$, $p < .001$, $\eta_p^2 = .21$) and a significant main effect of degrees of exaggeration ($F(2,94) = 71.1$, $p < .001$, $\eta_p^2 = .42$). There was also a significant interaction between these factors ($F(2,94) = 8.9$, $p < .001$, $\eta_p^2 = .08$). Pairwise comparisons using Bonferroni-corrected *t*-tests within the Proximal goal condition showed significant differences between non-exaggerated (i.e., Normal) and both exaggerated movements (Exaggerated: $t(94) = -3.7$, $p = .001$, $d = 0.76$; Very exaggerated: $t(94) = -5.6$, $p < .001$, $d = 1.3$). In the Distal goal condition, all pairwise comparisons between degrees of exaggeration yielded significant differences (all $t(94) < -4.9$, $p < .001$, $d > 1.1$). These results show that the larger the differences in movement velocities, the more consistently participants mapped them to the respective target location.

Mapping score. To investigate whether participants were more likely to map movements to targets in line with the motor-iconic prediction, we computed separate Bonferroni-corrected one-sample *t*-tests comparing the Mapping scores of each condition to 0 (i.e., random mapping direction). Mapping scores are displayed in Figure 4, where each dot represents an individual participant's Mapping score for each degree of exaggeration. We found that in the Proximal goal condition participants' responses were significantly different from chance (all $t(24) > 9.7$, $p < .001$, $d > 1.95$), showing a clear preference for Iconic mappings; see Figure 4A. In the Distal goal condition, however, participants' responses did not differ significantly from chance (all $t(23) > 0.21$, $p > .08$, $d > 0.04$). Thus, in the Distal goal condition, individual participants overall used a consistent mapping (resulting in a Consistency score that significantly differed from chance, as reported above), yet, across participants, there was no complete conformity as to the direction of that mapping (i.e., whether to map faster movements to the far target location and slower movements to the near target location or vice versa).

We conducted a 2x3 ANOVA with Mapping scores as dependent variable, goal type as between-subject variable and degrees of exaggeration as within-subject variable. The ANOVA yielded a significant main effect of goal type ($F(1,47) = 26.7$, $p < .001$, $\eta_p^2 = .28$), as well as a significant main effect of degrees of exaggeration ($F(2,94) = 15.5$, $p < .001$, $\eta_p^2 = .09$). The interaction between these two factors, however, was not significant ($F(2,94) = 0.1$, $p = .86$, $\eta_p^2 < .001$). Bonferroni-corrected post-hoc *t*-tests were used to analyze the main effect of exaggeration. In the Proximal goal condition, only the comparison between Normal and Very exaggerated yielded a significant result ($t(94) = -3.6$, $p = .001$, $d = 1.3$), whereas in the Distal goal condition the comparison between Normal and both exaggerated movements yielded significant results (Exaggerated: $t(94) = -3.0$, $p = .009$, $d = 0.6$; Very exaggerated: $t(94) = -4.0$, $p < .001$, $d = 0.6$).

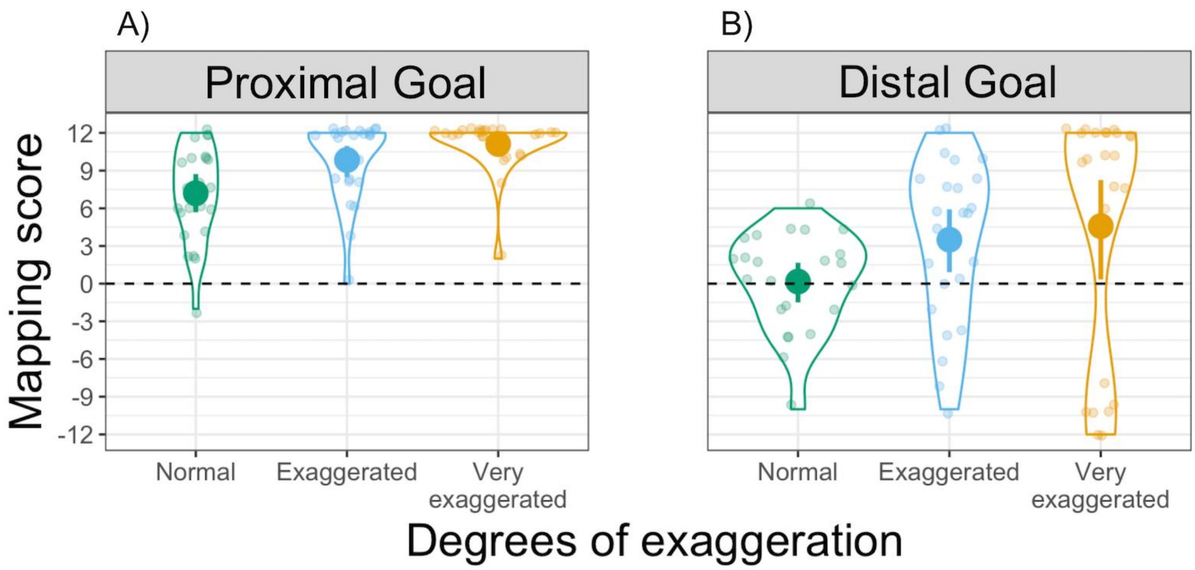

Figure 4. Mapping scores in Experiment 1

Distribution of Mapping scores in the (**A**) Proximal and (**B**) Distal goal conditions, for the three degrees of exaggeration. Each dot represents an individual participant, with one Mapping score for each degree of exaggeration. Violin plots represent the overall distribution of Mapping scores for each degree of exaggeration. The dashed horizontal line indicates random mapping direction.

2.2.3 Discussion

We hypothesized that participants would detect differences in the velocity of the observed movements and, based on these differences, consistently map the movements to one of the two target locations, particularly when the velocity differences were exaggerated. Our results support this hypothesis: Consistency scores significantly differed from chance and they were higher for more exaggerated movements.

In line with previous findings showing that movement velocity and distance are systematically related in natural aiming movements (Jeannerod, 1984), we had further hypothesized that participants would map slower movements to near target locations and faster movements to far target locations. Since these mappings preserve the relationship between movements and their most likely goals, they can be understood as “motor-iconic”. Our results in the Proximal goal condition provide evidence in support of this hypothesis, as shown by participants’ preference for Iconic mappings in this condition. The extent to which participants produced these mappings was also strongly influenced by the degree of exaggeration, as more exaggerated movements led the majority of participants to produce more Iconic mappings.

Our interpretation of this finding in terms of motor-iconicity implies that participants directly 'perceive’ the observed movements in terms of their proximal goals and that it is not a learned mapping. However, we need to acknowledge that the structure of our familiarization could, in principle, have contributed to the strong preference towards motor-iconic mappings in the Proximal condition. Participants were familiarized with two fully visible movements, each directed at one of the two target locations, and these two movements contained actual differences in their peak velocities. That means that, although both movements were Normal (i.e., non-exaggerated), the movement with the slightly higher peak velocity was directed at the far target and the movement with the slightly slower peak velocity was directed at the near target (see Figure 2A). This was a natural consequence of our decision to present real movements during familiarization (and not, e.g., movements artificially made equal). However, given that the familiarization only contained two trials that introduced participants to the details of the overall procedure, we consider it unlikely that this created a strong bias in participants’ responses.

Altogether, the findings of the Proximal goal condition are consistent with previous research showing that observers can derive useful anticipatory information from partially occluded actions and use this information to derive their partner's proximal goals in a joint setting (Vesper & Richardson, 2014).

Our results in the Distal goal condition indicate that movement modulations can also be used to extract information about an upcoming, more distal goal. Our findings in this condition show that, even when movements are not exaggerated, participants still produce a higher than chance rate of consistent movement-to-location mappings (see Consistency score, Figure 3B). However, when looking at the direction of these mappings (see Mapping score, Figure 4B), we found that there was no clear preference towards either of the two potential mapping directions in this condition: independent of exaggeration, about half of the participants chose to map faster movements onto near target locations and slower movements onto far target locations.

More generally, our pattern of results also seems to suggest that, depending on whether the observed movement reaches a proximal or a distal goal, participants might be resorting to different strategies to interpret the movements they see. When the movement attains only a proximal goal, as in the Proximal goal condition, participants need to observe the moving box and simulate its more likely goal given its velocity before occlusion (Prinz & Rapinett, 2008; Sparenberg et al., 2012). In such circumstances, participants can readily identify the motor-iconic relationship connecting the movements and the target locations, i.e., faster/slower movements leading to farther/nearer targets. When the movement achieves a distal goal, as in the Distal goal condition, participants cannot simply extrapolate the observed sliding movement towards the targets, as the box stops moving when it reaches the gray target in the middle of the screen. This is then followed by a sudden disappearance of the box from the display. As a consequence of this, the underlying motor-iconic relationship between the movement and its distal goal may not have been recognized in the Distal goal condition, as shown by the fact that participants, collectively, did not display a clear preference for Iconic over Non-Iconic mappings. To what extent is this lack of a general preference due to the spatiotemporal separation between a movement and a distal goal, or specifically to the fact that such separation was introduced in the present study by means of a sudden teleportation, is at present not clear. However, when asked to describe the strategy they used to solve the task, none of our participants in the Distal goal condition mentioned anything about the teleportation, while most of them made some reference to the difference in the velocity of the proximal movement, either in ways that are consistent with a motor-iconic interpretation (e.g., *My strategy depended on the speed of the box. Slow for near, fast for far*), or with its reversal (e.g., *If it looked like the box moved fast I selected NEAR*). This seems to suggest that participants were indeed able to interpret the relationship between the movements and their distal goals, despite the fact that the achievement of this latter goal was made possible by means of a teleportation of the box.

An alternative to the above interpretations could be that participants simply had problems to perceptually distinguish the velocity differences in the Distal goal condition. Although the high Consistency scores suggest that participants were able to discriminate between fast and slow movements, we conducted Experiment 2 to safely exclude the possibility that the pattern of results found in the Distal goal condition is due to difficulties in perceptually discriminating the different movement velocities.

2.3 Experiment 2. Perceiving velocity modulations

To determine whether the pattern of results found in the Distal goal condition really reflects an uncertainty about how to map the perceived velocity differences onto the occluded target locations, or whether it is simply due to difficulties in perceptually discriminating movements of different velocity, we conducted Experiment 2. Participants were shown the same animations as in Experiment 1, but were now asked to determine whether the movements were fast or slow. We predicted that participants would be able to discriminate the movements and that discrimination performance would be better for more exaggerated movements.

2.3.1 Methods

Participants. We recruited 49 participants (21 women; Age: *M* = 28.7 years; *SD* = 8.2 years) through Testable. The conditions of recruitment were identical to Experiment 1.

Stimuli, Design, & Procedure. Participants were presented with exactly the same animated movements as in Experiment 1, but their task was now to identify whether the movements were fast (by pressing the "f" key) or slow (by pressing the "s" key). As in Experiment 1, roughly half of the participants took part in the Proximal goal condition, the other half in the Distal goal condition. The only difference to Experiment 1 concerned the familiarization, where participants saw the occluded scene right away, and consequently never saw the two target locations on the right side of the screen. This choice was made to have participants focus on the velocity differences without making implicit associations about movement distance. As in Experiment 1, participants did not receive any kind of accuracy feedback.

2.3.2 Results

Data preparation. From participants' individual responses, we counted the total number of correct and incorrect responses for each movement, depending on the degree of exaggeration and the type of goal. We then subtracted these two values to obtain a *Discriminability score* that ranged from +12 (fully correct discrimination) to -12 (fully incorrect discrimination), which we could use to directly compare the results of Experiment 2 with the Mapping scores of Experiment 1.

Movement discrimination. The Discriminability score enabled us to measure the extent to which participants were able to correctly identify and categorize the movements as slow or fast on the basis of their differences in velocity. As expected, participants were able to correctly discriminate the movements, as shown by the significant difference from chance (i.e., higher than 0, see Figure 5) when movements were Normal, Exaggerated or Very exaggerated in both goal type conditions (all $t(23) > 5.1$, $p < .001$, $d > 1.0$, one-tailed).

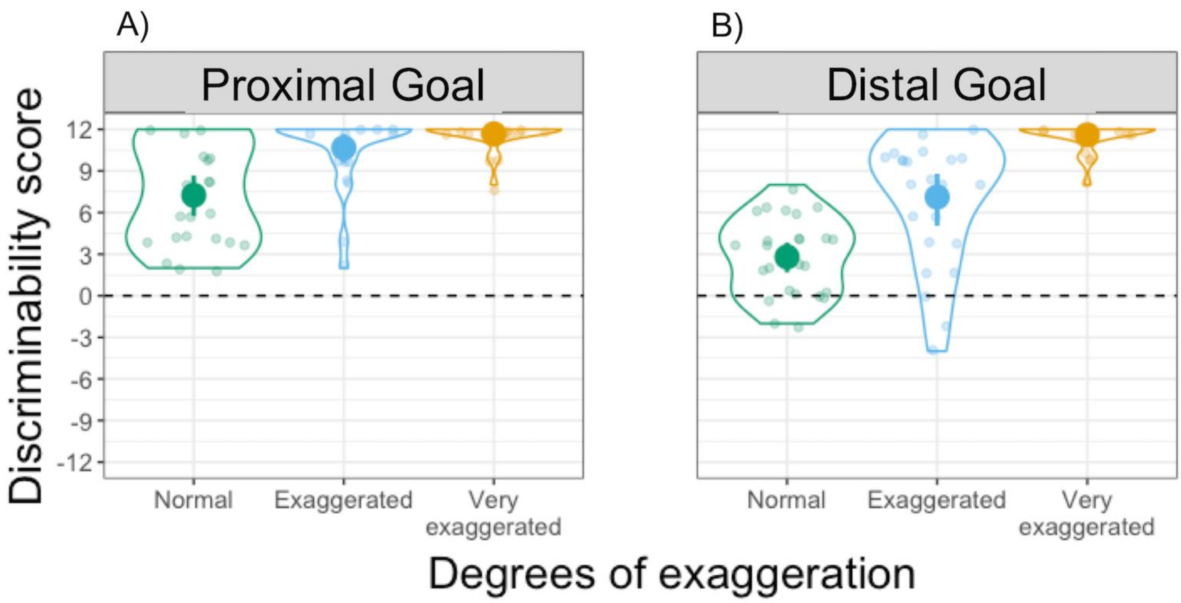

Figure 5. Discriminability scores in Experiment 2

Distribution of Discriminability scores in the (**A**) Proximal goal and the (**B**) Distal goal conditions, for the three degrees of exaggeration. The dashed line indicates chance discriminability (i.e., that movement velocities are not discriminable).

Comparison across experiments. To test whether the task (mapping different velocities to different target distances vs. discriminating different velocities) had an impact on participants' behavior, we conducted a 2x2x3 ANOVA comparing the Mapping scores of Experiment 1 and the Discriminability scores of Experiment 2 in both goal type conditions and across all three levels of exaggeration. The ANOVA yielded significant main effects of Experiment ($F(1,94) = 12.5$, $p < .001$, $\eta_p^2 = .08$), goal type ($F(1,94) = 45.2$, $p < .001$, $\eta_p^2 = .2$), and degrees of exaggeration ($F(2,188) = 65.3$, $p < .001$, $\eta_p^2 = .2$). We also found significant interactions between Experiment and goal type ($F(1,94) = 8.1$, $p = .005$, $\eta_p^2 = .05$), Experiment and degrees of exaggeration ($F(2,188) = 3.3$, $p = .04$, $\eta_p^2 = .01$) and goal type and degrees of exaggeration ($F(2,188) = 3.4$, $p = .04$, $\eta_p^2 = .01$). To further explore the main effects, we conducted Bonferroni-corrected post-hoc *t*-tests comparing participants' Mapping and Discriminability scores across the two Experiments for each degree of exaggeration. The analyses revealed significant differences between participants' responses across the two Experiments in the Distal goal condition, regardless of the degree of exaggeration (all $t(209) > -5.6$, $p > .04$, $d > 0.6$). In the Proximal goal condition, however, none of the pairwise comparisons across Experiments yielded a significant result (all $t(209) > -0.6$, $p > .5$, $d < 0.3$).

2.3.3 Discussion

The results of Experiment 2 indicate that participants can discriminate between the different velocities at a higher than chance level in both goal conditions. Along with the high Consistency scores of Experiment 1, these findings suggest that participants in the Distal goal condition are able to correctly distinguish the different movement velocities, and that their relatively lower Consistency scores and lack of preference for a unique mapping direction in Experiment 1 are not due to difficulties in perceptual discrimination. Instead, it seems that modulations in velocity, even when they are correctly categorized as fast or slow, are not uniformly associated to a unique distal goal in this condition.

Taken together, the results of the two experiments raise the possibility that some participants in the Distal goal condition in Experiment 1 might not have been aware of the underlying motor-iconic relation between the proximal movement and the distal goal. Specifically, participants who opted for Non-iconic mappings might not have been influenced by the intrinsic link between movement velocity and target distance (Jeannerod, 1984) that characterizes human natural performance. In order to address this hypothesis directly, we conducted Experiment 3.

2.4 Experiment 3. Interpreting goal type

Participants in Experiment 3 were presented with either Proximal goal trials followed by Distal goal trials, or vice-versa. We reasoned that being presented with trials where the movement attains its proximal goal first might subsequently help participants to recognize the underlying motor-iconic relation between velocity and distal goal. If that was the case, then we should expect participants who are initially presented with Proximal goal trials, followed by Distal goal trials, to apply the underlying connection recognized during the Proximal goal trials also to the Distal goal trials. We also presented another group of participants with Distal goal trials followed by Proximal goal trials, and reasoned that those participants who disregard the motor-iconic relation during the Distal goal trials might also disregard it later on, during the Proximal goal trials. This would suggest that establishing a particular interpretation of the movements early on might have the effect of overruling the effects of the motor-iconic relation.

2.4.1 Methods

Participants. We recruited 100 participants (35 women; Age: M = 29.6 years; SD = 9.7 years) through Testable. The conditions of recruitment were identical to Experiments 1 and 2.

Stimuli, Design, & Procedure. Participants were presented with the same animated movements as in Experiment 1 and their task was, as in Experiment 1, to choose the likely target location for each movement. This time, however, participants were randomly assigned to one of four conditions. The Proximal goal and Distal goal conditions were identical to the Proximal goal and Distal goal conditions of Experiment 1. The only difference pertained to the instructions as explained below. The other two conditions were a mix: Participants saw either three blocks of Proximal goal followed by three blocks of Distal goal trials (i.e., PG-to-DG condition), or three blocks of Distal goal followed by three blocks of Proximal goal (i.e., DG-to-PG condition).

During familiarization, participants were informed that the experiment consisted of two parts, and that their task would be to guess to which target location the box was being delivered. Halfway through the experiment, participants were introduced and familiarized with the second part. For participants in the Proximal goal and Distal goal conditions, this second familiarization was identical to the one they saw at the beginning. For those in the PG-to-DG and the DG-to-PG conditions, the new familiarization introduced the new layout, corresponding to what participants would see during the second part of the study (i.e., Distal goal in the PG-to-DG condition and Proximal goal in the DG-to-PG condition). In all conditions, participants were told that the movements they would see during the second part of the experiment had been recorded from another participant than those during the first part.

2.4.2 Results

***Data preparation*.** As in Experiment 1, we computed Consistency and Mapping scores for each participant across all three degrees of exaggeration. In order to render the data comparable across stable and mixed conditions, we computed separate scores for the first and second half of the experiment. As a consequence, the Consistency scores for each participant now ranged from 0 to 6 (Figure 6), and the Mapping scores ranged from -6 to +6 (Figure 7). Six participants who pressed the same key ten times in a row were excluded from the analyses.

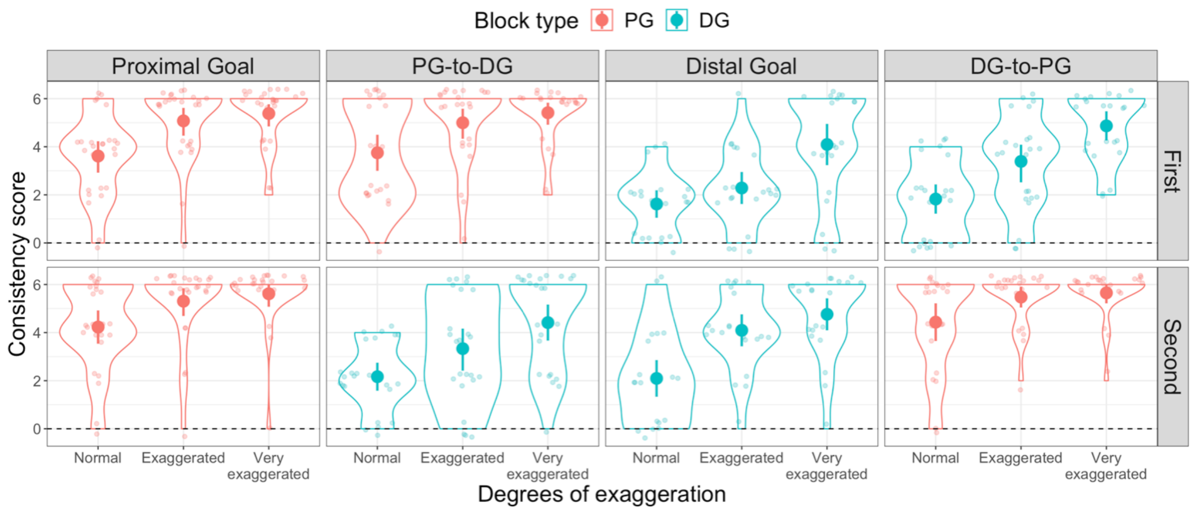

Figure 6. Consistency scores in Experiment 3

Distribution of Consistency scores in the (from left to right panel) Proximal goal, PG-to-DG, Distal goal and DG-to-PG condition. The upper panels display the distribution in the first half of the Experiment, while the lower panels display the distribution in the second half of the Experiment, for all three degrees of exaggeration. Violin plots in red represent Proximal goal trials, while violin plots in light blue represent Distal goal trials. The dashed line indicates random mapping (i.e., no consistency).

Mapping consistency and direction. We found that participants were able to produce consistent mappings across all degrees of exaggeration and goal type, as shown by the significant differences from a non-consistent baseline (i.e. 0) (all $t(20) > 4.9$, $p < .001$, $d > 1$, one-tailed) in both the first and second half of the experiment (see Figure 6). Regarding the direction of the mappings, we found that participants in the stable Distal goal condition were, collectively, equally likely to produce either Iconic or Non-Iconic mappings, as indicated by the non-significant difference from chance in that condition across all three degrees of exaggeration (all $t(20) > -1.0$, $p = 1$, $d > 0.02$, two-tailed), thus replicating our results from Experiment 1. Surprisingly, we did not find this pattern of results in the first half of the DG-to-PG condition, where participants were more likely to produce Iconic mappings (all $t(22) > 4.4$, $p < .004$, $d > 0.93$, two-tailed). This preference for Iconic mappings was also present in most other conditions (all $t(23) > 7.7$, $p > .001$, $d > 1.5$). The only two exceptions to this pattern of results were the previously mentioned Distal goal condition and the second half of the PG-to-DG condition in which our group of participants failed to display a clear preference towards any of the two mapping directions when the movements were either Normal ($t(23) = 2$, $p = 1$, $d = .4$) or Very exaggerated ($t(23) = 3.3$, $p = .06$, $d = .07$) (see Figure 7).

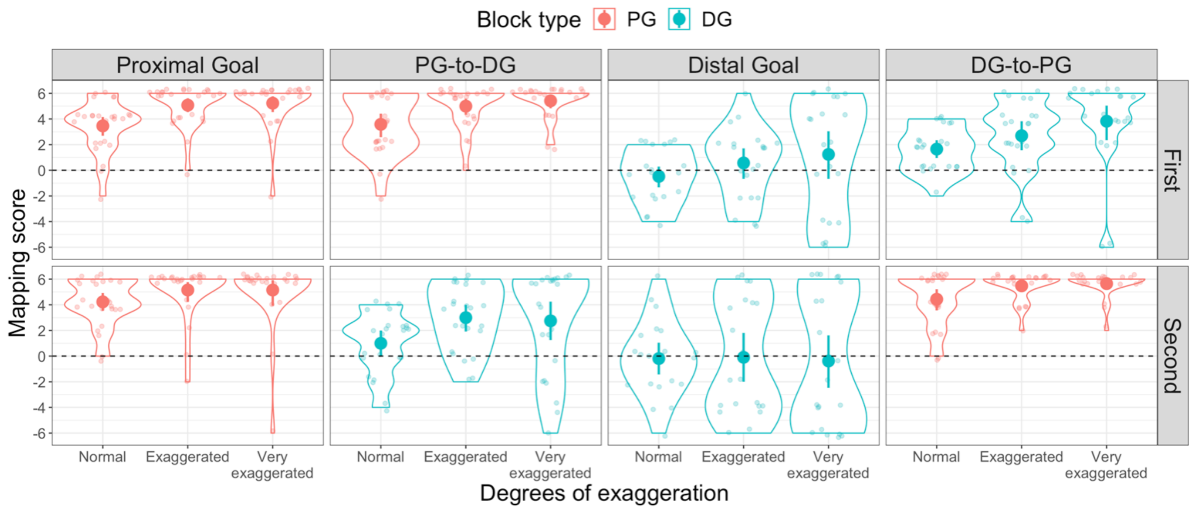

Figure 7. Mapping scores in Experiment 3

Distribution of Mapping scores in the (from left to right panel) Proximal goal, PG-to-DG, Distal goal and DG-to-PG condition. The upper panels display the distribution of Mapping scores of the first half of the Experiment, while the lower panels display the distribution of Mapping scores of the second half of the Experiment, for all three degrees of exaggeration. Violin plots in red represent trials in which participants were presented with Proximal goal trials, while violin plots in light blue represent trials in which participants were presented with Distal goal trials. The dashed line indicates random mapping direction.

Effect of previously viewed trials. To assess the effects of having been presented with a Distal or Proximal goal on participants' Mapping scores during the second half of the experiment we computed two ANOVAs. The first ANOVA compared the Mapping scores of the second half of the stable Proximal goal condition to those of the second half of the DG-to-PG condition. The ANOVA only yielded a main effect of degrees of exaggeration ($F(2,94) =$

10.6, $p < .001$, $\eta_p^2 = .07$), but no main effect of condition ($F(1,47) = .59$, $p = .45$, $\eta_p^2 = .008$). The second ANOVA compared the Mapping scores of the second half of the stable Distal goal condition to those of the second half of the PG-to-DG condition. The ANOVA yielded only a main effect of condition ($F(1,43) = 7.4$, $p = .009$, $\eta_p^2 = .1$). Bonferroni-corrected post-hoc *t*-tests revealed that Mapping scores were significantly higher in the second half of the PG-to-DG condition, compared to the second half of the Distal goal condition, but only when movements were exaggerated (Exaggerated: $t(85) = 2.8$, $p = .006$, $d = .8$; Very exaggerated: $t(85) = 2.8$, $p = .005$, $d = .6$).

Distal goal condition across experiments. In order to gain further insight into why participants' Mapping scores differ between the first half of the DG-to-PG condition and the first half of the Distal goal condition we compared these two conditions to the first half of the Distal goal condition of Experiment 1 using Bonferroni-corrected *t*-tests. The results revealed that, when comparing the Mapping scores of the Distal goal condition of Experiment 1 to those of the Distal goal condition of Experiment 3, none of the tests yielded significant differences across any of the three degrees of exaggeration (all $t(146) > 0.4$, $p > .7$, $d < 0.37$). No significant differences were found when comparing the first half of the No overlap condition of Experiment 1 to the first half of the DG-to-PG condition of Experiment 3 (all $t(146) > -1.8$, $p > .2$, $d < 0.77$) either.

2.4.3 Discussion

Experiment 3 showed that, as we predicted, being first presented with a situation in which the movement achieves a proximal goal (i.e., Proximal goal trials), helps participants recognize the underlying connection between movement velocity and distance, and that, once they recognize it, they can apply it to a situation in which the movement achieves a distal goal (i.e., Distal goal trials). This seems to suggest that the pattern of results found in Experiment

1, where some participants in the Distal goal condition preferred to use Non-Iconic mappings, is at least partially due to the fact that participants in that condition were not taking into consideration the underlying, motor-iconic relation that connects velocity to spatial locations. In Experiment 3 we addressed this hypothesis by assigning participants to a condition in which the connection between movement and target location was made explicit during the first half of the experiment, the PG-to-DG condition, and predicted that this would have an effect on participants' responses, specifically on their Mapping scores, during the second half of the experiment. The results in this condition support this hypothesis, as participants were more likely to produce Iconic mappings in the latter part of the experiment, right after being presented with trials in which the connection between the movements and their target locations was made explicit.

Our results also suggest that, although the explicit connection between movement and target location that was established thanks to trials in which the movements achieved a proximal goal has an effect on participant's responses, this precedence effect does not hold when we reversed the order of the trials, i.e., in the DG-to-PG condition. Specifically, participants in this condition, who were first presented with trials in which movements achieved a distal goal, still produced Iconic mappings consistently during the second half of the experiment (i.e., during the Proximal goal trials) in a way that was almost indistinguishable from that found in the Proximal goal condition.

In sum, our results in Experiment 3 point to the possibility that observing a movement that achieves a distal goal might be introducing, across participants, a relatively weak preference for any of the two possible mappings (i.e., Iconic or Non-Iconic). This can be seen in the lack of a general mapping preference in the Distal goal condition of Experiments 1 and 3. Such lack of a general preference for a specific mapping might also explain the differences we found between the first half of the DG-to-PG condition and the Distal goal condition.

Although identical during the first half of the experiment, these two conditions surprisingly led to differences in participant's Mapping scores. As our analyses suggest, this pattern of results might be simply due to the fact that in these two conditions our group of participants was either more likely than average to lack a preference towards any of the two mapping directions (i.e., the Distal goal condition of Experiment 3) or had a stronger preference than average to produce Iconic mappings (i.e., the first half of the DG-to-PG condition), or a combination of the two. In sum, this supports the idea that observing a movement achieving a distal goal weakens the preference for a particular mapping direction among some participants. This, in turn, might explain the differences found across the Distal goal and DG-to-PG conditions.

2.5 General Discussion

The aim of the present study was to investigate whether observers can interpret proximal communicative actions in terms of their distal goals. We hypothesized that participants would be able to detect communicative modulations of movement velocity and consistently interpret them in terms of the actions' distal goals (i.e., target locations), even though those distal goals could never be directly observed.

Our findings support this hypothesis, providing first evidence that observers can derive information about both proximal and distal goals from simple, one-dimensional movements. Specifically, participants in Experiment 1 were able to infer the likely proximal and distal goals of an action by relying on differences in movement velocity. In this respect, our findings are consistent with previous computational work suggesting that modulations of early kinematic features can be used strategically to communicate and disambiguate between an actor's distal goals (Donnarumma et al., 2017). Additionally, our results show that participants benefited from exaggerated velocity differences, allowing them to produce more consistent mappings. From this perspective, the present study provides further support to the previously established finding in SMC that observers can predict their partner's upcoming actions by relying on subtle kinematic modulations in their goal-directed movements (McEllin, Knoblich, et al., 2018; Pezzulo & Dindo, 2011; Sacheli et al., 2013; Vesper, Schmitz, et al., 2017; Vesper & Richardson, 2014), while at the same time making a novel contribution by extending these findings to a setting where observers need to infer an actor's distal goals.

Moreover, our results provide a first empirical demonstration that the way observers interpret communicative modulations is in part affected by whether the observed movement achieves a proximal goal only, or achieves a distal goal by means of attaining a more proximal goal first. When the movement achieves a proximal goal, observers display a clear preference towards interpreting the communicative modulations of that same movement in a motor-iconic manner, that is, in a manner that preserves the underlying relation between the movement and its likely goal. Although not completely absent, this preference for a motor-iconic relation is reduced when the movement achieves a distal goal. This might be partially due to the fact that participants fail to see the connection between a proximal movement and its more distal goal, especially when the transition between movement and goal is established indirectly by means of an unexpected event (a teleportation, as is the case in the present study). Understanding the factors that increase or decrease the preference for motor-iconic mappings when movements provide information about distal goals remains an important question for future research.

One hypothesis for why reference to distal goals reduces the preference for motor iconic-mappings is that seeing movements that achieve distal goals might induce a different, possibly more arbitrary or symbolic interpretation of the communicative modulations than seeing movements directed to proximal goals. Within the context of SMC, arbitrary mappings have been reported previously in studies in which participants were asked to coordinate their actions in tasks in which "Leader" participants with task-relevant information about specific target locations could opt to communicate this information to their naïve partners by means of either exaggerating the kinematics of their goal-directed movements (e.g., movement duration), thus resulting in what we here refer to as motor-iconic mappings, or alternatively by creating stable associations between non-dynamic features of these same movements (e.g., the dwell time on a target location) and the different target locations (Vesper, Schmitz, et al., 2017). These latter associations were described as "symbolic", as there seems to be no underlying motor relationship between the dwell time on a target location and the relative distance of that same target location. Interestingly, by switching from a more iconic to a more symbolic form of communication, actors were also able to create a temporal and functional separation between the instrumental (i.e. moving towards the target) and communicative (i.e. informing their partner) aspects of their movements, which in turn provided a more efficient way of informing their partners about their goals. Our Distal goal condition can be understood as an extension of these previous results, in that the achievement of a distal goal by means of attaining a more proximal one first may have also been interpreted by participants as a separation between the more instrumental aspect of the movement (i.e. delivering the box towards one of the target locations) and a more communicative one (i.e. informing an observer about a correct target location). As a consequence, the finding that participants displayed different mapping preferences in the Distal goal condition may be partly due to the fact that they took the first sliding movement to be a purely communicative movement, akin to a communicative gesture like pointing, with no explicit connection to an instrumental goal, let alone a distal one.

Relatedly, the strong preference for motor-iconic mappings in the Proximal goal condition raise the question of whether participants might be deriving the relationship between movement velocity and distance in a way that resembles an indexical relation, rather than a motor-iconic one. Indices are often described in semiotic theory as signs that carry information by virtue of them having an intimate relation with the objects they indicate, either in terms of spatio-temporal contiguity (e.g., a pointing gesture) or in terms of causality (e.g., smoke as a sign of fire) (C. S. Peirce, 1955). As we noted previously, there is a lawful relationship between the velocity and the distance of aiming movements during natural (i.e., non-exaggerated) performance, such that proximal movements directed at far target locations are performed with higher peak velocities (Jeannerod, 1984). This same intimate relation between velocities and distance was found in the way observers infer proximal goals from communicative movement modulations, as shown in the Proximal goal condition. Thus, it is a possibility that participants in that condition simply extracted some form of causal regularity or spatiotemporal contiguity that connects movement velocity and distance, thereby making velocity an indexical sign for proximal goals. From this perspective, introducing an explicit spatial and temporal separation between a proximal movement and its goal, as we did in the Distal goal condition, may have concurrently led to a change in the way participants derived information from the movements they observed, from a purely indexical interpretation to a more iconic one. This shift in the relation between signs and referents is similar to the one described by Keller, who illustrates this effect with the example of a real and a simulated yawn (Keller, 1998). Whereas a real yawn is simply taken as a sign caused by an underlying physiological state (e.g., tiredness) and is therefore an index, a simulated yawn is not, since it lacks, by definition, such causal antecedent. Instead, simulated yawns, because they are exaggerated simulations of the real yawn but still resemble it in some relevant way, are better described as icons. Similarly, movements that are directed at achieving distal goals but whose kinematic features resemble in some informative way those directed at proximal ones might also be taken as the iconic extension of the more basic, indexical causal relationship that connects proximal movements to their proximal goals. Under this view, an iconic interpretation is at least partly grounded on more basic indexical one, and therefore might explain the strong precedence effect that observing movements achieving their proximal goal has on subsequent interpretation of movements achieving a distal goal, as we found in Experiment 3.

The above discussion also points to the possibility that participants in the Distal goal condition may have thought of the sliding movement towards the middle of the screen as the first movement component of an implied two-step action sequence. Although participants in the Distal goal condition were only presented with a single sliding box movement, followed by the sudden disappearance of the box, we cannot discard the possibility that they might have tried to substitute the perceptual gap introduced by this sudden "teleportation" by simulating a second sliding movement, similar to the one they had just seen (Prinz & Rapinett, 2008; Sparenberg et al., 2012). Whether participants did indeed generate such simulations, and whether these simulations affected the way they mapped the movements onto the target locations, are both open questions for further research.

Given our initial interest in addressing questions related to the understanding of communicative modulations, our study focused exclusively on the receiver end of the interaction, that is, on whether and how observers interpret proximal communicative actions in terms of their proximal and distal goals. Moreover, as the movements participants saw in the present study were created for the purpose of the experiment, the question remains whether participants who had the task to actively inform someone else using our setup would display a spontaneous preference towards creating mappings that preserve the motor-iconic relation between proximal movements and distal goals (i.e., Iconic mappings), or would rather opt to reverse this mapping (i.e., Non-Iconic mappings). Alternatively, communicators might opt to convey such information through other relevant movement parameters, either continuous ones as when people directly modulate the total duration of their movements while trying to keep the velocity constant, or discrete ones such as movement pauses or sudden changes in movement direction. These different strategies are not a mere theoretical possibility, since they correspond to what some participants in our studies reported at the end of the experiment when they were asked which communicative strategy they would choose. While the majority described strategies that were consistent with a motor-iconic one (e.g., "*I'd move it very slowly for the near location and very quickly for the far location*") or with its reversal (e.g., *I would move the box slower to reach the further location*), a few did mention other movement parameters (e.g., jerk: "*I would move it jerkily – one jerk for near and two for far*"; deceleration: "*Ending the placement of [the box] faster for far* […] *slower for near*.").

A final consideration concerns the role of objects used in order to convey information about action goals. Our study presented participants with animations of a moving box with a cursor attached to it and a context in which this box was smoothly "delivered" along a black line towards one of two target locations. By doing this, we tried as much as possible to highlight the relevant differences in the velocity of the box, while also minimizing other factors, such as the weight or fragility of the box, that could have had an influence on participant's interpretations. For example, one would normally expect a heavy or fragile object to be moved slowly in order to maintain control during its transport and placement. If applied to our task, this would mean that an actor would be forced to adopt a slower pace when sliding the box, thus having a direct effect on the options of communicative strategies at her disposal. Again, future studies could explore the role that objects, with their specific physical properties and affordances, might play in the way communicators flexibly adapt their modulations to provide information to observers (Schmitz et al., 2018b).

The present study offers valuable perspectives for future research on joint action and communication. As argued at the outset, it is likely that people engaged in a joint action will try to predict their partner's distal goals by relying on a wide variety of kinematic cues. Being able to make such long-term predictions can be particularly useful in situations where co-actors produce complex action sequences that require the coordination of actions at different temporal levels (Schmitz et al., 2018a), such as dancing or playing football. In such scenarios, providing relevant information about the upcoming distal goals early on in the action sequence would be a useful and effective manner to facilitate coordination.

2.6 Supplementary material

2.6.1 Movement recording and exaggeration

Spontaneous goal-directed movements were collected by the first author using an in-house PsychoPy script that recorded mouse movements continuously within a setup that looked identical to the layout of the Proximal goal condition in Experiment 1. The animated box was attached to the mouse cursor. Cursor movements were constrained by locking them to the horizontal axis, resulting in one-dimensional sliding movements. Additionally, the animated box could only move in one direction, from the left side of the screen towards the green targets on the right.

To avoid any biases in the collection of these movements, our in-house script was set to randomly select trials to near and far targets, until a nearly equal number of at least 50 near and 50 far movements were recorded. This procedure led to a total of 105 movements (see Figure 1). We then smoothened each individual movement and averaged all near and all far movements, respectively, thus obtaining two non-exaggerated movements, one for each target location (henceforth “Normal near” and “Normal far” movements). This averaging procedure was key, as it allowed us to identify systematic differences between near and far movements (e.g., peak velocity), while also controlling for more subtle differences between individual movements (e.g., in jitter). Importantly, the averaging procedure preserved distinguishable human-like features (e.g., bell-shaped velocity profile, with a faster initial phase and slower final phase) which have been frequently reported in studies looking at rapid aiming movements (Jeannerod, 1984).

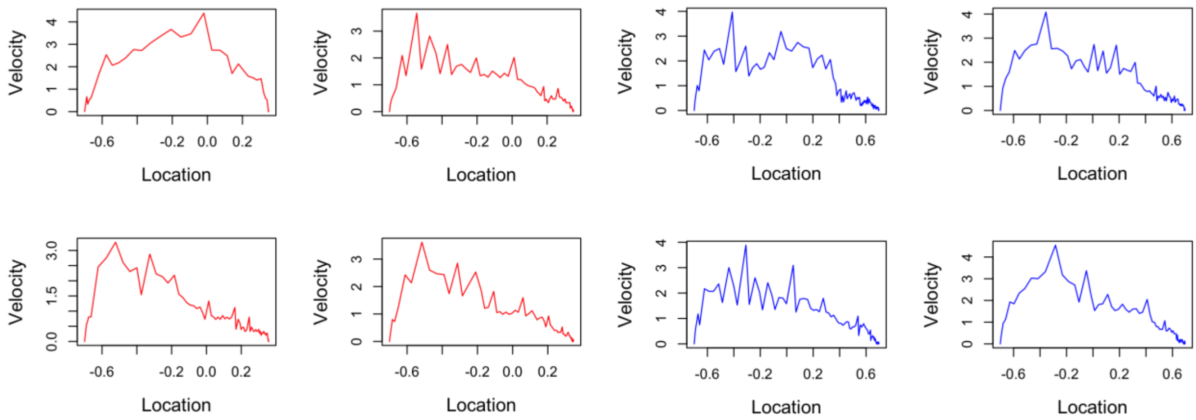

Figure S1. Unsmoothed velocity profiles of sliding movements

Subset of raw velocity profiles of sliding movements recorded, before averaging and smoothening. Movements directed to the near target are colored, while those directed to the far target are colored in blue.

Exaggerated movements were generated in three steps. First, we computed the standard deviation of near and far movements separately. Then, we identified the highest value (peak velocity) for each averaged (i.e. Normal near and Normal far) movement. Finally, we rescaled both of these movements by either subtracting one and two standard deviations from the peak velocity of the Normal near movement, or by adding one and two standard deviations to the peak velocity of the Normal far movement.

2.6.2 Movement rescaling in Distal goal condition

For the Distal goal condition the six Normal and Exaggerated movements were reshaped so that their endpoints would all converge towards the middle of the screen. To do so, we used the "rescale" function in R Studio to manually specify the maximum values of the location vectors of each movement.

The rescaling procedure yielded velocity profiles that, unsurprisingly, differed in average velocity from the original velocity profiles in the Proximal goal condition. This is simply due to the fact that the change of the movement endpoints via rescaling implies that the new velocity profiles will be either compressed (when the endpoint is shifted closer to the movement origin) or expanded (when the endpoint is shifted away from the movement origin). As a consequence of this, the difference in peak velocity between the two non-exaggerated (i.e. Normal) movements in this condition was inverted with respect to the original movements in the Proximal goal condition. Specifically, while in the Proximal goal condition the Normal movement directed towards the far target (i.e. Normal far) had a higher peak velocity than the one directed towards the near target (i.e. Normal near), this relationship was reversed in the Distal goal condition, where the Normal movement directed to the near target (i.e. Normal near) became slightly faster than the one directed to the far target (i.e. Normal far). Since we hypothesized that participants would try to produce consistent mappings based on the velocity of the movements, we decided to account for this specific reversal in the velocity of the two Normal movements in our analyses of the Distal goal condition.

2.6.3 Data preparation in Distal goal condition

As we pointed out above, the rescaling procedure led to a reversal in the velocity profiles of the two Normal movements in the Distal goal condition, such that the originally faster movement (i.e. Normal far) became slightly slower than the originally slower one (i.e. Normal near). To account for this reversal in our analyses, we decided to code "f" key responses for Normal near movements as Iconic, and "n" key responses for Normal far movements as Non-iconic in this condition. We applied the same conversion with the Normal far movements (i.e. "n" key responses for Normal far movements were coded as Iconic, while "f" key responses for Normal far movements were coded as Non-iconic). This is in line with our hypothesis that faster movements will be more likely to be mapped to far target locations, while slower ones will be mapped to near target locations (Jeannerod, 1984).

Chapter 3. Simulations underlying the prediction of distal goals

3.1 Introduction

Many social interactions rely on our capacity to predict other people's actions and to quickly adapt our behavior accordingly (Knoblich & Sebanz, 2006). Even a simple social act like shaking hands requires one to anticipate and monitor the other person's arm and hand movements so as to make one's palm meet the exact same spot on the other person's hand – all this in a matter of just a few seconds (Melnyk et al., 2014). Such simple joint actions, and also more complex ones such as playing a piano duet, benefit from (and sometimes are even made possible by) our capacity to predict the outcome and timing of others' actions while these unfold (Wilson & Knoblich, 2005), or even prior to their initiation (Kilner et al., 2004). In order to facilitate such predictions, interacting agents will often resort to a variety of behavioral strategies, aptly known as "coordination smoothers" due to their role in simplifying coordination demands during interaction (Vesper et al., 2010; Clark, 1996). One such smoother involves the modulation of certain kinematic parameters of an instrumental action in order to make the action more salient and readable to an observer, who can then predict the action outcome more easily. For example, one of the pianists in the duo might lift her arms with a high amplitude right before starting to play, thus providing more explicit anticipatory information about the timing of her immediate actions to her co-performer (Goebl & Palmer, 2009). Crucially, such modulations are characterized by the fact that their underlying goals have a "dual nature" (Pezzulo et al., 2019, p. 3): the person performing the action can achieve simultaneously an instrumental goal (e.g., pressing one of the keys on the piano) and a communicative goal (e.g., informing the co-performer about the exact moment when she will start playing) (Clark, 2005). In keeping with previous research showing how such communicative modulations can be used to facilitate coordination (Pezzulo et al., 2013), we will refer to these as "sensorimotor communication" (henceforth 'SMC').

Most previous research on SMC has focused on how actors coordinate their actions by relying on communicative modulations of instrumental actions directed to proximal goals (see Pezzulo et al., 2019 for a review). Proximal goals can be understood as those goals whose achievement is the result of a single transitive or intransitive movement, like placing an object on a table or making a step to a new position in space, respectively. Proximal goals are therefore directly tied, both temporally and spatially, to the movements leading to their immediate achievement. Distal goals, on the other hand, are achieved by two or more successive movements, each with its own proximal sub-goal. Consequently, and unlike proximal goals, the early movement(s) preceding the achievement of a distal goal are both spatially and temporally separated from it by one or more intermediate movements. A well-studied example of such distal goals, and one that will be the focus of the present work, are the goals that result from a two-step action sequence, as when a grasping movement towards a piece of food (i.e. a first movement component) is followed by bringing it to one's mouth (i.e., a second movement component) (Haggard, 1998; Marteniuk et al., 1987).

The present set of studies draws on two lines of research: first, research on people's production and understanding of actions directed at distal goals, and second, research on SMC. Based on these two lines of research, we ask under which conditions observers can predict the distal goal of a two-step action sequence when observing communicative modulations of a first movement component. Before addressing this question, we will first review relevant findings in the motor control literature suggesting that acting towards distal goals leads to changes in kinematic features of early movement components (Gentilucci et al., 1997). These changes in movement kinematic can be used by observers to predict distal goals of observed actions (Lewkowicz et al., 2013). Then, we will argue that such predictive processes play a key role in

SMC, where observers use communicative modulations of instrumental actions to simulate distal goals, thereby establishing what we call “motor-iconic” relations between modulations in the kinematics of early action components and the action’s distal goal.

3.1.1 Distal goals in action production and observation

A large body of research in motor control has demonstrated that the way we plan and execute simple motor acts in non-communicative contexts is highly sensitive to our upcoming actions and to the distal goals achieved by these actions (Jeannerod, 1984). For example, a natural reaching movement towards an object is performed with different velocities depending on whether the reach is followed by a careful placing of the object or by throwing the object in a large box (Marteniuk et al., 1987; Armbrüster & Spijkers, 2006). Similar effects of distal goals on early action kinematics have been found when participants are asked to reach towards an object and then place it in a target location that varies in size, position or relative distance (Gentilucci et al., 1997). Specifically, in the latter study, reaching velocity was higher and maximal finger aperture was larger when the object had to be placed in a far target relative to a near one (see also Rand et al., 1997; Johnson-Frey et al., 2004). Furthermore, studies on the “end-state comfort effect” report a tendency for participants to grasp objects in bio-mechanically awkward ways in order to ensure a comfortable position at the end of their movements (D. A. Rosenbaum et al., 1990; Cohen & Rosenbaum, 2004), thus again suggesting a strong influence of distal goals on the planning and execution of early movement components.

One way researchers have sought to explain how distal goals affect early action kinematics is by appealing to the mediating role of motor or action representations (Jeannerod, 1994, 1997; Butterfill & Sinigaglia, 2014). These representations play a key role in guiding actions towards their goals while the action unfolds (D. A. Rosenbaum et al., 1992; Haggard, 1998). As a consequence, performing an action directed at a distal goal, like throwing an object into a basket after picking it up, can sometimes lead to visible changes in the kinematics of the early movement components (in this case, the grasping of the object) leading towards such goal (for a similar proposal in terms of "coupled planning" see Lewkowicz & Delevoye-Turrell, 2020). Other features of distal goals, such as their expected value (i.e., reward), can also lead to visible changes in early movements and, in turn, increase people's motor performance (Adkins & Lee, 2021; Galaro et al., 2019).

Besides their role in guiding the performance of one's own actions, motor or action representations are also involved in understanding and predicting other people's instrumental actions and goals (Pacherie & Dokic, 2006; Sinigaglia & Butterfill, 2016, 2022). Crucially, these predictions often go beyond the mere online anticipation of an unfolding movement or its proximal (i.e., immediate) goal, as they can extend to the prediction of more distal goals (Lewkowicz et al., 2013). For example, research suggests that motor regions in the brain are differently activated depending on whether an observed reaching movement towards a piece of food is then followed by the distal goal of placing it in a container or of eating it (Fogassi et al., 2005). What makes these predictions possible is the fact that observers rely on early kinematic information contained in the movements (Manera, Becchio, Cavallo, et al., 2011; Cavallo et al., 2016; Lewkowicz et al., 2013; Koul et al., 2016) but also on objects placed in the immediate vicinity of the agent (Bach et al., 2014; Cattaneo et al., 2009).

Most research on action understanding has shown that observers can use kinematic information as a cue to infer information about the observed actor's (both proximal and distal) goals. This has recently led researchers in the field of joint action to ask whether, in social interactions, co-actors would not only rely on each other's kinematic information to coordinate their actions, but would also actively exaggerate their movements, thus facilitating the achievement of a joint goal by making their action goals easier to predict (Manera, Becchio, Cavallo, et al., 2011; McEllin, Knoblich, et al., 2018; Cavallo et al., 2016; Lewkowicz et al., 2015; Quesque et al., 2013; Georgiou et al., 2007; Sartori et al., 2011). These situations, in which co-actors actively inform each other about their goals by modulating the kinematics of their movements, are at the core of SMC. We turn to these now.

3.1.2 Proximal and distal goals in SMC

Studies on SMC have mostly focused on settings in which co-actors are required to coordinate their actions by, for example, aiming at one of three target locations either in synchrony (Vesper & Richardson, 2014) or sequentially (Vesper, Schmitz, et al., 2017). Typically, the design of these studies is such that the information allowing co-actors to achieve coordination, like the location of the correct target, is allocated to one of the actors only, the "Leader". Furthermore, participants are not allowed to speak nor gesture during the performance of these tasks. Instead, Leaders inform their naïve "Followers" by modulating key kinematic features of their instrumental actions. For example, Leaders have been shown to exaggerate the amplitude of their aiming movements in order to disambiguate between target locations (Vesper & Richardson, 2014), or to increase or decrease their wrist height as a way of informing the Follower about the part of the object they are about to grasp (Sacheli et al., 2013). These modulations are more easily perceived by Followers, who can use them to disambiguate between various proximal goals (McEllin, Sebanz, et al., 2018; Pezzulo & Dindo, 2011; Trujillo et al., 2018; Vesper & Richardson, 2014). Once they are able to detect and disambiguate between different proximal goals, Followers can adapt their behavior in a timely manner, thereby leading to successful temporal and/or spatial coordination (Vesper, Schmitz, et al., 2017; Vesper, Schmitz, Safra, et al., 2016). Thus, given that Leaders exaggerate their movements (in studies on SMC), and that these exaggerations are in turn used by Followers to facilitate their predictions of the Leader's goals, these movements have sometimes been described as "signals" (Pezzulo et al., 2019), to be contrasted with the natural, non-exaggerated movements in studies on action observation, where observers rely on these movements as "cues" to predict the goals of the agent.

At the core of SMC is the capacity for Leaders to make their proximal goals more discriminable by producing kinematic modulations, and for Followers to understand such modulations as conveying anticipatory information about these goals. However, as we noted previously, the predictions made during action observation are not limited to proximal goals, but are also directed at more distal goals. Thus, this raises the question of whether communicative modulations of early action components can be used by observers to predict an agent's distal goals.

Some of the findings on action observation we reviewed earlier provide preliminary evidence indicating that observers can derive information about distal goals when observing the initial stages of naturally performed reaching movements (Ansuini et al., 2016; Becchio et al., 2008). Thus, these studies indicate that observers can infer distal goals when observing instrumental movements that are not communicatively modulated by their actors. As far as we know, only two studies have directly tested whether observers can also do this for communicative modulations of actions. Donnarumma and colleagues (2017) conducted a series of computationally-guided analyses on the processes involved in the recognition of two-step action sequences. Their analyses suggest that a performing agent who actively modulates her first movement component so as to increase the similarity in its kinematic features with respect to a second movement component (by means of "coarticulation", see Fowler, 1980) can facilitate an observer's early recognition of a distal goal.

In our own previous study (2023) we asked directly whether observers would be able to understand kinematic modulations of a first movement in terms of distal goals. To do so, we presented participants with animations of a box sliding at different velocities towards an intermediate target location, after which the box was automatically delivered towards one of two final, occluded target locations, far or near. Participants were then asked to select the final location towards which they thought the box had been delivered. The findings indicated that participants benefit from modulations in velocity of the sliding movement to infer the distal goal. This was indicated by their capacity to consistently map the different movements onto different target locations. Specifically, when participants observed movements with higher peak velocity they were more likely to map them onto the "far" target location, whereas movements with lower peak velocity were more likely to be mapped onto the "near" target location.

We referred to this particular relation holding between modulations in movement velocity and distal goals as "motor-iconic", as it corresponds to a regular relationship between the velocity and distance of unconstrained aiming or grasping movements towards targets of varying distance in either one-step (Jeannerod, 1984) or two-step action sequences (Gentilucci et al., 1997). In these studies, participants tend to aim with higher velocity at far compared to near targets. As a consequence, we apply the notion of "motor-iconicity", which refers to the regular relationship between non-communicative movements and goals as found during natural performance, to the relationship between exaggerated movements and their goals in the context of SMC.

3.1.3 Two mechanisms for simulating a distal goal

Given that motor-iconic relations capture a regular relationship between communicative modulations of actions and their distal goals, it is unclear how such relationship is established when observing two-step action sequences. Based on the results of Dockendorff et al. (2023), we can conclude that such relationship can be captured by establishing a "direct" link between a first movement component and a distal goal. That is, a first movement component containing sufficient kinematic information can be used to predict an upcoming distal goal without relying on a second movement component. This implies that it is sufficient to observe modulations in the initial movement component to predict distal goals. Since such a process relies on linking kinematic information from an initial movement directly to its distal goal, we will refer to it as a "movement-to-goal" simulation (see Figure 1A).

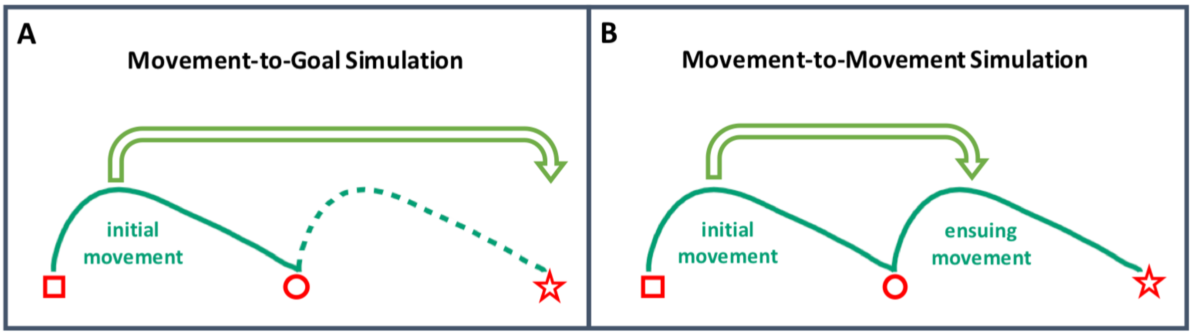

Figure 1. Two kinds of simulation enabling participants to link communicative modulations of an initial movement to a distal goal.

The green lines are a schematic representation of velocity profiles of sliding movements. The red square indicates the starting point of the initial movement, the red circle both the endpoint of the initial movement and the starting point of the ensuing movement, while the red star indicates the endpoint of the ensuing movement, i.e., the distal goal of the sequence. (**A**) In the movement-to-goal simulation, observing an initial movement is sufficient to predict the distal goal of the action, and can therefore bypass the simulation of the ensuing movement. (**B**) In the movement-to-movement simulation, the initial movement enables an observer to simulate an ensuing movement, which in turn enables the simulation of the distal goal.

Another possibility is that a second movement component within a sequence plays an active role in the process of linking communicative modulations of first movement components to distal goals. In such scenarios, the initial movement is fed into a simulation of an ensuing movement, rather than into a direct simulation of the distal goal. Since this process involves taking into account the mediating role of the second movement component in the sequence, we will refer to it as "movement-to-movement" simulation (see Figure 1B).

Note that we use the term "simulation" here to refer to the capacity observers have to link movements to distal goals. This process of linking movements and goals might rely on (online) motor mirroring, as proposed by research on action understanding and SMC (Kilner et al., 2004; Pezzulo et al., 2013). However, in the context of our experiment, it is also possible that participants built these relations based on other processes, such as action-effect associations (Hommel, 2009), generalized mechanisms of statistical learning (Ahlheim et al., 2014) or even on Gestalt-like principles applied to action recognition (Trujillo & Holler, 2023).

Besides the two types of simulation discussed above, it is also possible that observers do not engage in any type of simulation but that, once they identify that the movements contain some relevant kinematic information ("communicative signals"), they try to establish a stable link between them and the distal goals. This mapping is arbitrary in the sense that the observer makes a choice at the beginning of the experiment and does not engage in a simulation (and so does not necessarily choose a mapping that would be seen as motor-iconic). This form of arbitrary interpretation contrasts with a situation in which an observer disregards the communicative signals present in the movements altogether, in which case one would expect them to link the movements to their distal goals in a random fashion.

To address these different possibilities, we present three experiments in which participants needed to infer a distal goal on the basis of modulations in velocity and/or duration of a first movement within a two-step sequence. The purpose of Experiment 1 was to investigate how different kinematic information present in the first movement component of the sequence would be then fed into a simulation of a distal goal. Then, in Experiments 2 and 3 we address the question of whether participants interpret communicative modulations of a first movement component in a way that takes into account the second movement component. Thus, we were interested in determining whether participants rely on movement-to-goal simulations, or on movement-to-movement simulations (or on both).

3.2 Experiment 1. What kinematic information is needed for a simulation of a distal goal?

The aim of the first experiment was to identify how different kinematic information present in a first movement component helps participants to simulate its distal goal. To do so, we compared two-step action sequences in which the first movement component contained continuous velocity information (Sliding-Sliding condition) or only discrete duration information (Jumping-Sliding condition). The initial movement was then followed by a continuous movement towards an occluded, and therefore unknown, target location. If the presence of velocity and/or duration information facilitates the prediction of a distal goal, participants should consistently map fast initial movements onto far targets and slow initial movements onto near targets, consistent with previous findings in motor control (Gentilucci et al., 1997). Moreover, we expect participants to benefit from the exaggerations in velocity and/or duration, as such exaggerations generally facilitate the prediction of goals during SMC (Pezzulo et al., 2013).

3.2.1 Methods

Participants. We recruited 50 participants (29 women; Age: M = 30.8 years; SD = 9.6 years), 25 per condition, through the online testing platform Testable (https://www.testable.org/). All participants gave their informed written consent prior to inclusion in the study, in accordance with the Psychological Research Ethics Board (PREBO; reference number 2020_04). Sample size was determined using the Superpower statistical package (Lakens & Caldwell, 2021) on RStudio (R Core Team, 2013). The design and analyses of the study were preregistered on OSF, at https://osf.io/px96b. The videos used in all experiments are publicly accessible on OSF, at https://osf.io/pv74b/. Data collection for all studies was performed between May 2021 and November 2022.

Stimuli. The basic layout for both experimental conditions of Experiment 1 is shown in Figure 2. In both conditions, participants were first presented with a stationary box with a mouse cursor attached to it. The box and cursor were displayed within a black hexagonal location on the left-hand side of the screen. During familiarization, participants also saw a grey hexagonal intermediate location in the middle of the screen and two green hexagonal target locations on the right-hand side of the screen. A black horizontal line, along which the box moved during the trials, connected the initial location to the grey hexagonal area in the middle of the screen and to the two green target locations in both conditions (see Figure 2).

During trials, a black occluder covered different sections of the display. During the first movement component (i.e., when the box moved from the initial location to the grey intermediate target location), the occluder only covered the green target locations (Figure 2A). Right before the beginning of the second movement component (i.e., before the box started sliding from the intermediate target location towards the green target locations) the occluder was widened, thus covering the intermediate grey target and the box itself (Figure 2B). This prevented participants from seeing the initial acceleration phase of the second sliding movement, from which they could have derived information about where the box would then move (i.e., the near or far target).

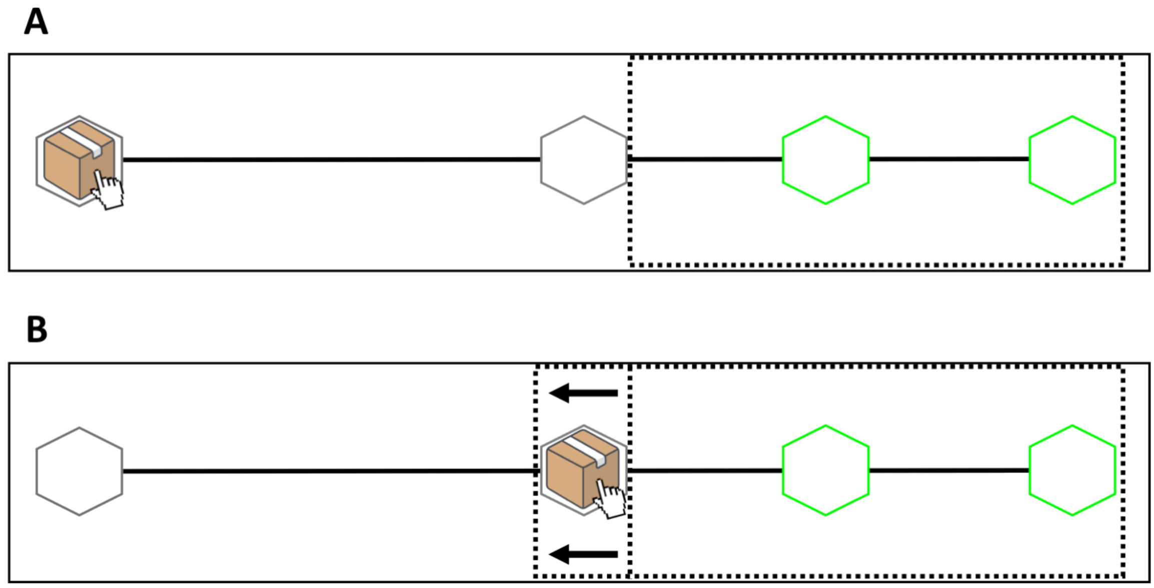

Figure 2. Experimental layout used in Experiment 1.

The black dotted line represents the outline of the occluded area in both conditions during trials. (**A**) At the beginning of the trial, the occluder only covers the green target locations, leaving visible the grey target location in the middle of the screen. (**B**) Once the box had arrived at the intermediate target location, and right before it started sliding towards the green targets, the occluder became wider in order to cover the beginning of the second sliding movement.

Design. The experiment consisted of a mixed 2x3 factorial design, with one between-subject factor (first movement type) and one within-subject factor (degrees of exaggeration). The between subject factor manipulated whether the movement of the box from the initial to the intermediate location was presented as a continuous, fully visible sliding movement between these two locations (i.e., Sliding-Sliding condition) or whether the same movement was presented as an invisible, discrete "jumping" movement from the origin to the intermediate location with a duration that matched the total duration of the corresponding sliding movements (i.e., Jumping-Sliding condition). The within-subject factor manipulated whether and to what degree the animated movements were exaggerated in terms of peak velocity or duration (i.e., Normal (i.e., no exaggeration), Exaggerated, Very exaggerated). The procedure used to create such movements is the same used in Dockendorff et al., (2023). Briefly, it consisted in exaggerating two non-exaggerated movements by adding or subtracting two predefined values from their peak velocities, thus yielding a total of six movements: the two original non-exaggerated ones (i.e., Normal), a slow and a fast one (i.e., Exaggerated), and a very slow and a very fast one (i.e., Very Exaggerated) (see Supplementary Material). Since participants in the Jumping-Sliding condition only saw a snapshot of the movement before onset and after offset, these exaggerations were never perceived as such. Instead, participants could only rely on the differences in total duration, corresponding to the time interval between the disappearance and reappearance of the box.

Procedure. During the familiarization, participants were first presented with the complete task layout (as illustrated in Figure 2), but without the occluder covering the green target locations. Participants then saw two successive Normal movements of the box towards the middle target, where it stayed stationary for approximately 1500 ms, after which the box started sliding towards either of the two target locations (order counterbalanced across participants). After seeing the two Normal movements, participants in both conditions were asked to select the target location where the box had moved by pressing the "n" key (for near) or "f" key (for far). In the Jumping-Sliding condition, participants were also told at the very beginning of the experiment that they would "only see a snapshot of the box disappearing at the starting position and then a snapshot of it reappearing in the middle grey target. Thus, you will not see the actual sliding movement performed by the previous participant, but only the beginning and end of it."

Next, a black occluder covered the target locations (Figure 2A) and participants were presented with two more Normal movements of the box towards the middle target (order counterbalanced). Right before sliding from the middle target to the (now occluded) target locations, the occluder became wider on the left side, thus covering the box. Participants were explicitly told that this change in occluder size indicated the onset of the second sliding movement (Figure 2B). Finally, participants were told that the movements they would see had been recorded from a previous participant, and that this participant had produced such movements "in ways that would help others guess to which green target location he/she was moving the box".

Experiment. Participants performed 36 experimental trials, divided into six blocks. In each trial, they were presented with an animation of the box either sliding along the black line in the Sliding-Sliding condition or disappearing from the origin and reappearing at the intermediate location in the Jumping-Sliding condition. Participants in both conditions were then prompted to answer to which target location they thought the box had been delivered. Note that in both conditions the second sliding movement was never visible to participants during the trials, but, as explicitly stated in the familiarization, its onset was indicated by the widening of the occluder. A trial was completed when participants pressed one of the two assigned keys ("n" or "f"), corresponding to either the "near" or "far" target locations. Each block contained all six degrees of exaggeration, presented in random order. Participants did not receive feedback about their performance at any point. At the end of the experiment, participants were asked to fill out a short questionnaire about their experience with the task.

3.2.2 Results

Data preparation. We categorized participants' responses as *Iconic* or *Non-Iconic* mappings. Iconic mappings refer to trials where participants pressed the "n" key in response to movements with lower peak velocity (and thus longer total duration) and the "f" key in response to movements with higher peak velocity (and thus shorter total duration). Non-Iconic mappings, on the other hand, refer to those trials where participants reversed this association, i.e., by pressing "f" in response to movements with lower peak velocity (i.e., longer total duration) and “n” in response to movements with higher peak velocity (i.e., shorter total duration).

Two dependent variables were computed from participants’ number of Iconic and Non-Iconic mappings. Calculating the *absolute* difference between the total number of Iconic mappings and the total number of Non-Iconic mappings, separately for each first movement type and each degree of exaggeration, yielded a *Consistency score* for each participant ranging from 0 to 12. A score of 0 meant that participants mapped velocities and/or durations randomly to targets and a score of 12 meant that participants mapped with absolute consistency. Calculating the *signed* difference between Iconic and Non-Iconic mappings gave us the *Mapping score*, which could range from +12 (fully iconic mappings) to -12 (fully non-iconic mappings). A Mapping score of 0 meant that participants had no preference for any of the two mapping directions.

Participants who pressed the same key (either “n” or “f”) at least ten times in a row were excluded from further analysis. Based on this criterion, 4 participants were excluded in Experiment 1.

Consistency score. To test whether participants benefited from the exaggerations by producing more consistent mappings, and whether such effect differed depending on whether velocity information was present in the movements, we conducted a 2x3 ANOVA with Consistency scores as dependent variable, first movement type (Sliding, Jumping) as between-subject factor and degrees of exaggeration (Normal, Exaggerated, Very exaggerated) as within-subject factor. The ANOVA yielded a main effect of degrees of exaggeration ($F(2,88) = 23.4$, $p < .001$, $\eta^2 = .19$), but no main effect of first movement type ($F(1,44) = 1.8$, $p = .18$, $\eta^2 = .023$). The interaction between these two factors was not significant ($F(2,88) = .36$, $p = .7$, $\eta^2 = .003$) (see Figure 3). This pattern of results suggests that regardless of whether the first movement contained velocity information (i.e., Sliding) or not (i.e., Jumping), participants were able to consistently map the different movements to the target locations (*M* Consistency Score: Sliding-Sliding = 5.21, Jumping-Sliding = 4.22) and benefited from their exaggerations.

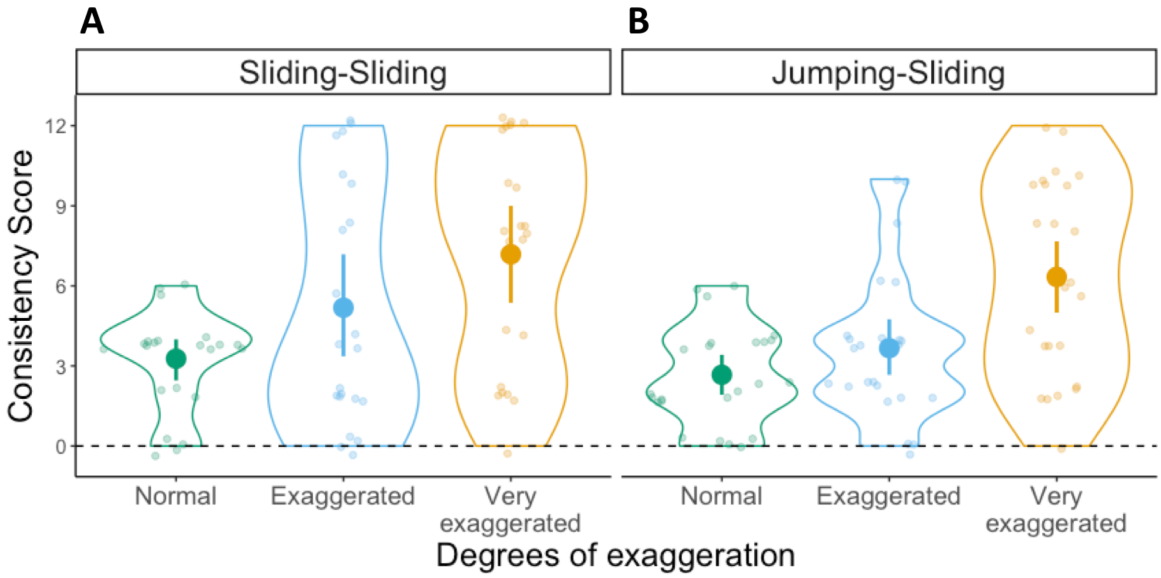

Figure 3. Consistency scores in Experiment 1.

Distribution of Consistency scores in the (**A**) Sliding-Sliding and (**B**) Jumping-Sliding conditions, for the three degrees of exaggeration. Each dot represents an individual participant, with one Consistency score for each degree of exaggeration: Normal in green, Exaggerated in light blue and Very exaggerated in yellow. Violin plots represent the overall distribution of Consistency scores for each degree of exaggeration. The dashed horizontal line indicates a hypothetical value for random mapping (i.e., no consistency).

Mapping scores. A 2x3 ANOVA was conducted with Mapping scores as dependent variable, first movement type as between-subject factor and degrees of exaggeration as within-subject factor. The ANOVA revealed a main effect of first movement type ($F(1,44) = 20.1$, $p < .001$, $\eta^2 = .22$), but no main effect of degrees of exaggeration ($F(2,88) = 2.3$, $p = .11$, $\eta^2 = .02$). Furthermore, we found a significant interaction between these two factors ($F(2,88) = 7.6$, $p < .001$, $\eta^2 = .06$). To further explore the main effect of first movement type we conducted Bonferroni-corrected *t*-tests comparing the Mapping score distributions for each degree of exaggeration across the two between-subject conditions (e.g., Exaggerated in Sliding-Sliding versus Exaggerated in Jumping-Sliding). The comparison of Mapping scores across conditions between Normal movements did not reach significance ($t(94) = 1.3$ $p = .18$, $d = 0.6$), but the one between Exaggerated ($t(94) = 4$, $p < .001$, $d = 1.2$) and Very Exaggerated movements ($t(94) = 5.3$, $p < 001$, $d = 1.2$) did (see Figure 4). This suggests that participants benefit from communicative modulations of peak velocity (i.e., Sliding-Sliding) since they are able to map the movements to the targets in a motor-iconic fashion. Modulations in movement duration (i.e., Jumping-Sliding), however, did not lead to more motor-iconic mappings.

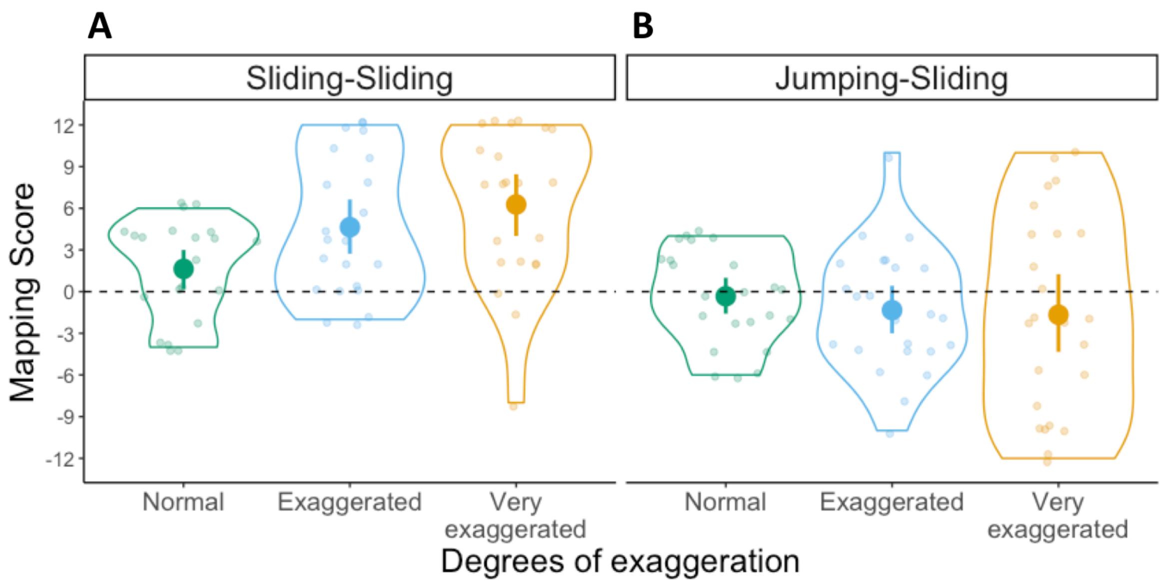

Figure 4. Mapping scores in Experiment 1.

Distribution of Mapping scores in the (A) Sliding-Sliding and (B) Jumping-Sliding conditions, for the three degrees of exaggeration. Each dot represents an individual participant, with one Mapping score for each degree of exaggeration. Violin plots represent the overall distribution of Mapping scores for each degree of exaggeration. The dashed horizontal line indicates random mapping direction.

3.2.3 Discussion

The general aim of Experiment 1 was to identify how different kinematic information present in a first movement component within a two-step action sequence is used by observers to infer its distal goal. We were particularly interested in determining whether the presence of continuous velocity information or of discrete duration information presented in the communicative modulation of the first movement of the sequence would enable participants to simulate the distal goal.

Participants in the Sliding-Sliding condition, who saw the box sliding with different peak velocities, were able to use the differences in velocity to simulate the distal goal of the action sequence, as indicated by the high consistency of their mappings (Figure 3) and the fact that such mappings were motor-iconic, that is, in line with the underlying regular relation between movements and goals found in natural movement performance (Figure 4) (Jeannerod, 1984; Gentilucci et al., 1997). This contrasted with the Jumping-Sliding condition, in which participants were also highly consistent in their responses, but less likely to produce motor-iconic mappings. This indicates that in both conditions, participants were able to create stable mappings based on the communicative modulations present in the movements, but that they relied on motor-iconic mappings only if the first movement contained continuous velocity information. When the initial movements only contained discrete information about their total duration, as in the Jumping-Sliding condition, this was less likely to trigger the appropriate simulation process that enables the prediction of the distal goal.

One possible explanation for the lower motor-iconic mappings in the Jumping-Sliding condition is that the lack of velocity information introduces, among some participants, a more arbitrary interpretation of the relationship between movements and distal goals. Since this arbitrary relation is the result of participants simply choosing a particular link between movement duration and target location, some of them chose a link that reversed the motor-iconic mapping, thus mapping faster peak velocities to the near target and lower peak velocities to the far target. Our results of the Jumping-Sliding condition provide support for this hypothesis, given that participants are able to map the movements onto the target locations consistently (see Figure 3B), but they fail, collectively, to display a unique mapping preference (see Figure 4B), since they were sometimes nearly split between motor-iconic mappings and their reversal (Motor-iconic: N = 10; Reversed: N = 13; see Very Exaggerated in Figure 4B).

Interestingly, what may have led participants to reverse the mapping in the Jumping-Sliding condition is an altogether different intuition about the relationship between movement duration and target distance. The intuition is nicely captured in the following description given by one of our participants in the Jumping-Sliding condition who, when asked to explain her strategy to solve the task, replied: “just assuming that [the] movement that *took longer* was supposed to indicate a *longer journey*” (italics ours). In other words, according to this intuition, longer movements (i.e., movements with longer total durations, and thus, in our task, with lower peak velocities) are more likely to travel longer distances than shorter movements (i.e., movements with shorter total durations, and thus, in our task, with higher peak velocities). As a consequence, this intuition leads to a reversal of the above specified motor-iconic mapping. This reversal has, interestingly, already been reported in previous studies on SMC in which participants used total duration to communicate information about target distance (Vesper, Schmitz, et al., 2017). We return to a more detailed discussion of this mapping reversal in the General Discussion.

One final explanation for the lower motor-iconic mappings in the Jumping-Sliding condition would be that participants simply fail to perceptually discriminate the differences in duration between the movements. However, the Consistency scores in that condition show that participants are able to create stable associations throughout the experiment between movement durations and target locations, regardless of how exaggerated they are. Participants could not form such stable associations if they were not able to discriminate the differences in movement duration (Dockendorff et al., 2023).

In sum, our findings indicate that observers rely on velocity information to infer and simulate a distal goal. However, this pattern is consistent with both types of simulation we propose (see Figure 1). Concretely, participants in the Sliding-Sliding condition may have either used the velocity of the initial movement to simulate the velocity of the upcoming one, and subsequently the distal goal (i.e., movement-to-movement simulation), or they may have disregarded the upcoming movement and simulated the distal goal directly based on the first movement only (i.e., movement-to-goal simulation). In order to gain a clearer understanding of which type of simulation is being used, we conducted Experiment 2.

3.3 Experiment 2. What kind of simulation underlies the prediction of a distal goal?

In Experiment 2, participants saw a first sliding or jumping movement to the intermediate location (as in Experiment 1), but this time were familiarized with a second jumping movement rather than a second sliding movement as in Experiment 1. This meant that the first movement (Sliding or Jumping) was now followed by a second movement that, unlike Experiment 1, did not contain velocity information, but only duration. Accordingly, we labeled the two conditions of the present experiment Sliding-Jumping and Jumping-Jumping.

If participants are engaging in movement-to-goal simulations, then we expect them to be able to simulate the distal goal only when the observed movement contains continuous velocity information, and regardless of whether the second movement contains velocity information or not. Thus, we expect participants in the Sliding-Jumping condition to map more consistently in the motor-iconic direction than in the Jumping-Jumping condition, and thereby produce a pattern of responses similar to the one found in the Sliding-Sliding condition of Experiment 1. In contrast, if participants engage in movement-to-movement simulations, which take into account the presence (even if implied) of a second continuous movement in the process of simulating the distal goal, we expect participants in the Sliding-Jumping condition to have difficulties in simulating the distal goal, and therefore produce less motor-iconic mappings than in the Sliding-Sliding condition of Experiment 1. More generally, we expected participants to be able to use the communicative modulations of movement velocity and/or duration, and therefore to benefit from their exaggerations by producing more consistent mappings.

3.3.1 Methods

Participants. We recruited 50 participants (25 women; Age: *M* = 29.8 years; *SD* = 9.9 years) through Testable. The conditions of recruitment were identical to Experiment 1. Based on the same exclusion criteria used in Experiment 1, we excluded 2 participants from our analyses. The design and analyses of this study were pre-registered on OSF (https://osf.io/upg92)

Stimuli. The layouts for each experimental condition of Experiment 2 were very similar to Experiment 1 (see Figure 5). The only difference with respect to Experiment 1 pertained to the size of the occluder, which in both conditions was kept the same size throughout the trial, thus leaving the intermediate target always visible.

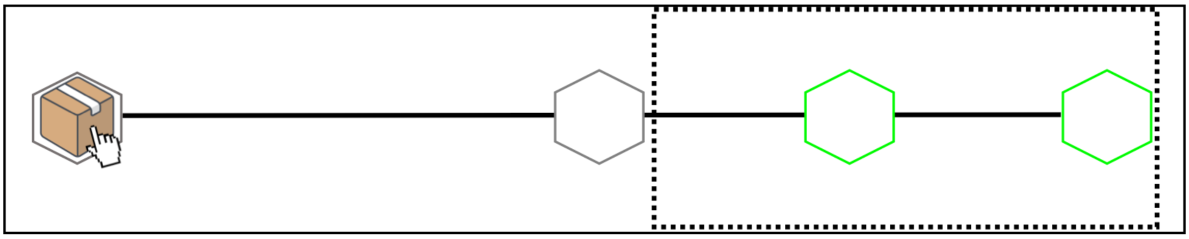

Figure 5. Experimental layout used in Experiment 2.

Experimental layout in the Sliding-Jumping and the Jumping-Jumping condition. The black dotted line represents the outline of the occluded area during trials. The (near and far) target locations are displayed in light green and the intermediate location is displayed in grey.

Design and Procedure. As in Experiment 1, Experiment 2 consisted of a 2x3 mixed-factorial design with first movement type as a between-subject factor and degrees of exaggeration as a within-subject factor.

Participants were familiarized with the full layout and were presented with two normal movements. Once the box reached the intermediate location, where it stayed stationary for approximately 1.5 secs, it then disappeared and reappeared in one of the two target locations (order counterbalanced across participants). Note that the duration of these second jumping movements matched the duration of the second sliding movements used in the familiarization of Experiment 1. Thus, the duration of the second jumping movement directed to the far target was 2.74 secs, while the one to the near lasted 2.23 secs. After seeing the two Normal movements, participants in both conditions were asked to select the target location where the box had moved by pressing the "n" key (for near) or "f" key (for far). Then, they were presented with two more Normal movements, but this time the green target locations were covered with the black occluder (Figure 5).

In each trial, participants were presented with an animation of the box either sliding along the black line in the Sliding-Jumping condition or disappearing from the origin and reappearing at the intermediate location in the Jumping-Jumping condition. Unlike Experiment 1, however, participants saw the box disappearing from the middle gray target, which indicated the beginning of the second jumping movement. Participants then selected the target location to which they thought the box was delivered by either pressing the "n" key (for near) or "f" key (for far).

3.3.2 Results

Consistency score. We conducted a 2x3 ANOVA with Consistency scores as dependent variable, first movement type as between-subject factor and degrees of exaggeration as within-subject factor. The ANOVA yielded a main effect of degrees of exaggeration ($F(2,92) = 16.8$, $p < .001$, $\eta^2 = .14$), and an interaction between degrees of exaggeration and first movement type ($F(2,92) = 3.6$, $p = .03$, $\eta^2 = .035$). However, there was no significant main effect of first movement type ($F(1,46) = 1.5$, $p = .22$, $\eta^2 = .03$). This pattern of results suggests that participants benefited from communicative modulations of movements containing continuous velocity information (i.e., Sliding-Jumping, $M = 5.77$), but not necessarily from modulations of movements containing only discrete duration information (i.e., Jumping-Jumping, $M = 4.82$) (see Figure 6).

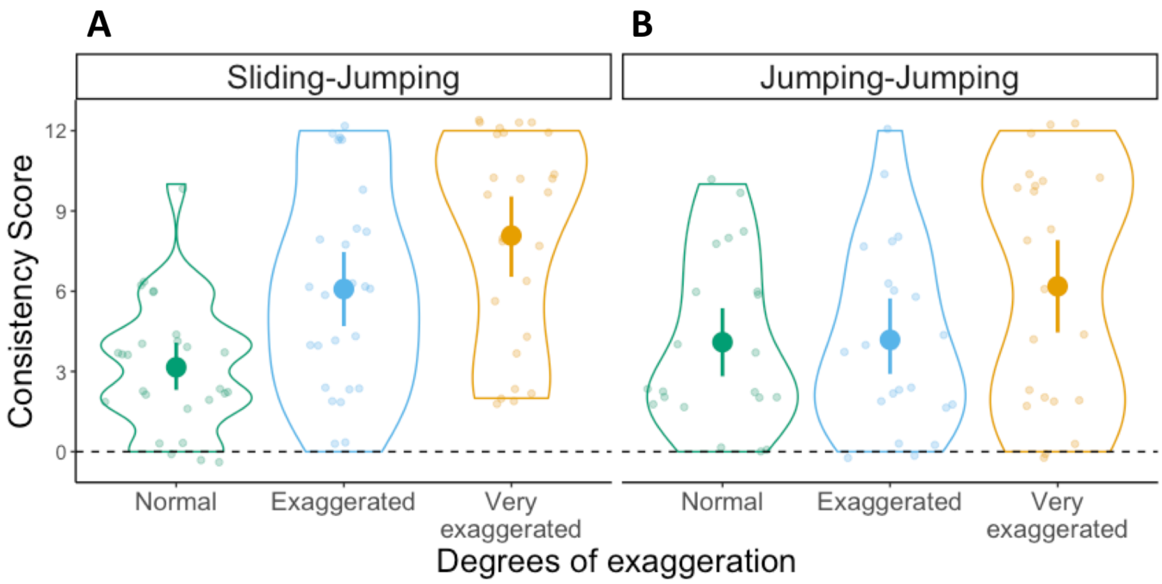

Figure 6. Consistency scores in Experiment 2.
Distribution of Consistency scores in the (A) Sliding-Jumping and (B) Jumping-Jumping conditions. Violin plots represent the overall distribution of Consistency scores for each degree of exaggeration. The dashed horizontal line indicates a hypothetical value for random mapping (i.e., no consistency).

Mapping score. We conducted a 2x3 ANOVA with first movement type as a between-subject factor and degrees of exaggeration as a within-subject factor. The ANOVA yielded a main effect of first movement type ($F(1,46) = 9.6$, $p = .003$, $\eta^2 = .12$) but not of degrees of exaggeration ($F(2,92) = 1.8$, $p = .16$, $\eta^2 = .01$). The interaction between these two factors did not reach significance ($F(2,92) = 2.1$, $p = .13$, $\eta^2 = .01$). This replicates our findings of Experiment 1 and confirms that participants are more likely to map movements motor-iconically if the first movement contains continuous velocity information rather than discrete duration one (see Figure 7).

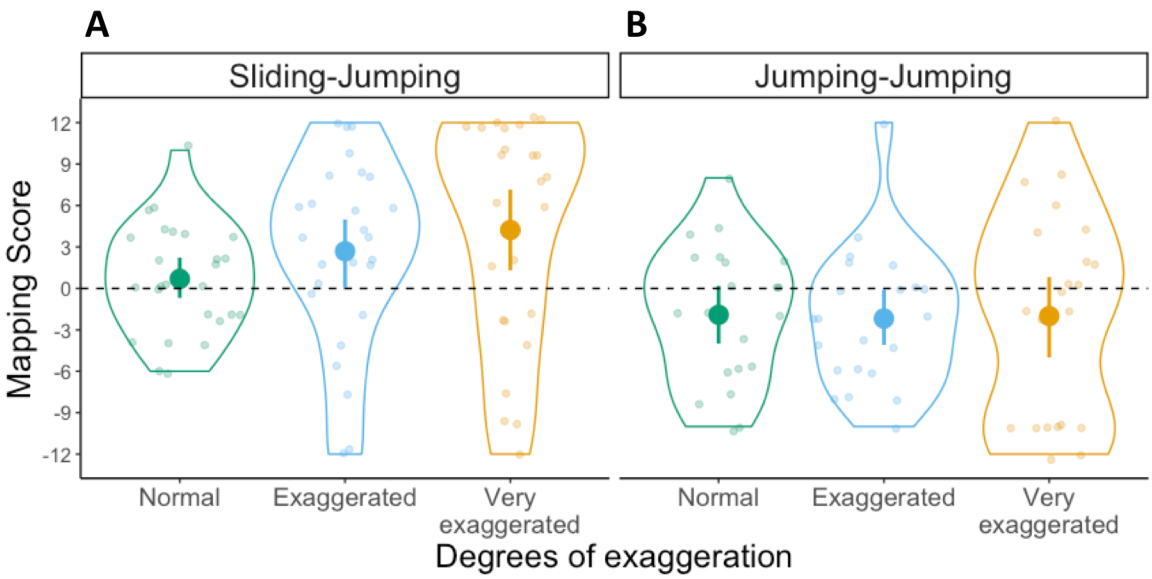

Figure 7. Mapping scores in Experiment 2.

Distribution of Mapping scores in the (A) Sliding-Jumping and (B) Jumping-Jumping conditions, for the three degrees of exaggeration. Violin plots represent the overall distribution of Mapping scores for each degree of exaggeration. The dashed horizontal line indicates random mapping direction.

Type of Simulation. To see whether the presence of a second movement that contains velocity information plays a role in simulating a distal goal, we conducted a between-Experiment comparison. Specifically, we conducted a 2x3 ANOVA using the Mapping scores of the two conditions in which participants saw velocity information in the first movement component (i.e. a Sliding movement: *Sliding*-Sliding in Exp. 1 and *Sliding*-Jumping in Exp. 2) as a dependent variable. Thus, the between-subject factor in the ANOVA was the second movement type, which in this case either contained velocity information (Sliding-*Sliding* in Exp. 1) or not (Sliding-*Jumping* in Exp. 2). As in our previous analyses, the within-subject factor was the degrees of exaggeration. The ANOVA yielded a main effect of degrees of exaggeration ($F(2,92) = 12.5$, $p < .001$, $\eta^2 = .08$), but no main effect of second movement type ($F(1,46) = 1.44$, $p = .23$, $\eta^2 = .02$) and no significant interaction ($F(2,92) = .27$, $p = .76$, $\eta^2 = .002$).

3.3.3 Discussion

In Experiment 1 we found that participants were able to rely on communicative modulations of continuous velocity information to simulate the distal goal of the action sequence. When participants only received discrete information about the duration of the movement, they failed collectively to display any strong mapping preference, thus suggesting that at least some participants did not simulate a distal goal with such minimal duration information (or at least not in the way we predicted, i.e., using motor-iconic mappings). These findings, however, do not allow us to draw conclusions about the type of simulation that observers were using to predict the distal goal on the basis of velocity information. To gain more knowledge about these underlying simulation processes, we modified the second movement component in both conditions of Experiment 2, from a continuous movement to a discrete one that did not contain velocity information. This created a situation in which movement-to-movement simulations of a distal goal on the basis of observing an initial movement containing velocity information were made more difficult.

The results of Experiment 2 show that, despite the fact that participants' simulations in the Sliding-Jumping were made more difficult (because there was no continuous second movement to map onto), they were still able to predict the distal goal from observing modulations in movement velocity in that condition. This was indicated by the higher number of motor-iconic mappings and the increasing consistency (as a function of degrees of exaggeration) in the Sliding-Jumping condition compared to the Jumping-Jumping condition. In this latter condition, unexpectedly, participants did not benefit from the exaggeration of movement duration. These results therefore provide a first indication that observers can predict distal goals directly by engaging in movement-to-goal simulations.

A second indication that observers engage in movement-to-goal simulations, and thus might not need to rely on an implied second movement in order to predict a distal goal, is the fact that participants in the Sliding-Jumping condition collectively agreed as much on the motor-iconic mappings as those in the Sliding-Sliding condition of Experiment 1. Thus, this suggests that it is sufficient for observers to observe velocity modulations to simulate its distal goal, and that a second movement does neither interfere nor contribute to such capacity.

However, when looked at more descriptively, our results also point to differences in how some participants interpret the movements when they rely on direct movement-to-goal simulations. Indeed, the mapping scores in the Sliding-Jumping condition of Experiment 2 show that while the majority of participants mapped the movements onto the target locations motor-iconically ($N = 19$), a small number reversed this mapping ($N = 6$), thus suggesting that those participants either arbitrarily chose a mapping at the beginning of the experiment or, as we suggested in our Discussion of Experiment 1, had opposing intuitions about the mapping.

Finally, a comparison between Sliding-Jumping in Experiment 2 with the Sliding-Sliding condition of Experiment 1 showed no significant difference in performance. This leaves open the possibility that the velocity information in the second movement might not always have an effect on participant's capacity to simulate a distal goal (see Figure 4 and 7). In other words, this result is consistent with both direct movement-to-goal simulations and movement-to-movement simulations. In sum, the findings of Experiment 1 and 2 suggest that observers rely on movement-to-goal simulations to link modulations in movement velocity to a distal goal, but that movement-to-movement simulations could in some circumstances also play a role. To test this possibility more directly, and thus to see if there are circumstances in which observers rely on the second movement to simulate a distal goal, we conducted Experiment 3.

3.4 Experiment 3. When do movement-to-movement simulations enable the prediction of a distal goal?

To explore more directly under which circumstances movement-to-movement simulations facilitate motor-iconic interpretations of communicative modulations, we conducted Experiment 3 in which we reversed the direction of the first sliding movement while we manipulated, across two conditions, whether the second movement contained continuous velocity information or not. Unlike the two previous experiments in which both movements approached the target locations and had therefore a similar connection to the distal goal, Experiment 3 creates an asymmetry between the two sequential movement components such that the movement directions are opposed and only the second movement is directed towards the distal goal. With this manipulation, we had two aims. First, we wanted to test whether participants rely exclusively on the velocity of a movement, regardless of its direction, when simulating the distal goal. Second, by reversing the first movement, we aimed at weakening its connection to the goal and, in contrast, to highlight the connection of the second one with respect to the goal. Based on this, we expected participants to be more likely to rely on movement-to-movement simulations in this experiment.

We predicted that if participants engage in movement-to-movement simulations, then they should be able to simulate the distal goal in the Sliding-Sliding condition, where they are presented with a second movement containing velocity information that they can feed into their simulation, but less so in the Sliding-Jumping condition, where this second movement containing velocity information is absent. However, if participants disregard the second movement and only engage in movement-to-goal simulations, then we do not expect a difference across the two conditions, as only observing the first movement should enable them to simulate the distal goal. Finally, since the reversed movements in the present study still preserved differences in peak velocity, we predicted that in both conditions participants would benefit from the exaggerations by producing more consistent motor-iconic mappings.

3.4.1 Methods

Participants. We recruited 50 participants (11 women; Age: *M* = 31 years; *SD* = 12.2 years) through Testable. The conditions of recruitment were identical to Experiment 1 and 2. We excluded one participant from our analyses. The design and analyses of this study were pre-registered on OSF (https://osf.io/3g24f)

Stimuli. The full layout without the black occluder for both conditions is presented in Figure 8A. Participants saw a layout in which the box was initially displayed in the middle of the screen, at the rightmost end of the black line connecting the middle of the screen to the grey hexagonal location on the left-hand side of the screen.

During trials, once the box was moved from the starting location towards the grey location on the left, the black line along which the box had moved either got stretched out, thereby connecting the grey location to the green target locations on the right side of the screen (in the Sliding-Sliding condition, see Figure 8B) or completely disappeared from the display (in the Sliding-Jumping condition, see Figure 8C).

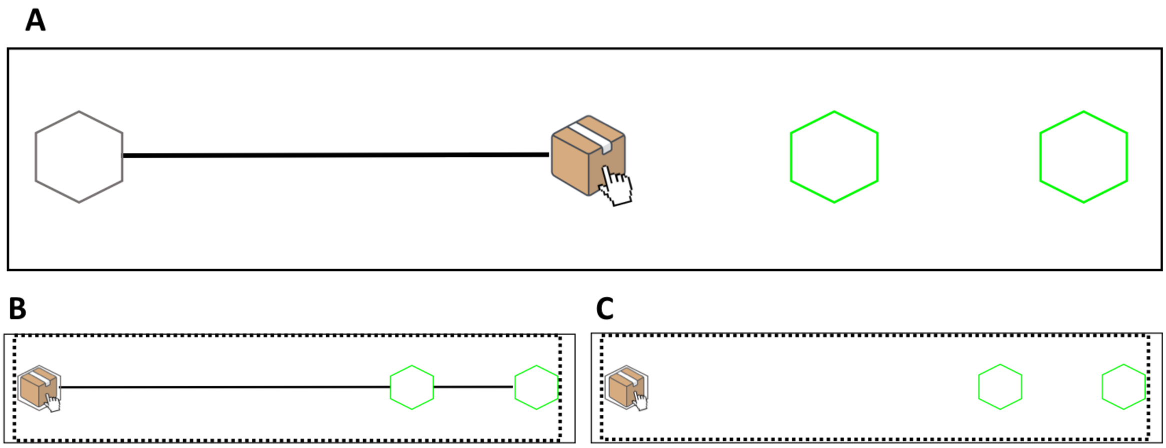

Figure 8. Experimental layout used in Experiment 3.
(A) Participants in both conditions were first present with the full layout, without the occluder. Once the box was moved towards the intermediate location on the left, the black line changed differently depending on the condition. (B) In the Sliding-Sliding condition, the black line connected the intermediate location to the green target locations. (C) In the Sliding-Jumping condition, the black line disappeared from the display.

Design and Procedure. Experiment 3 consisted of a mixed-factorial design with second movement type as a between-subject factor and degrees of exaggeration as a within-subject factor.

As in Experiments 1 and 2, participants were first presented with the full task layout, without the occluder (Figure 8A). In Experiment 3, however, the box moved continuously from the starting location now located in the middle of the screen, towards the intermediate target now located on the left. Participants saw two Normal sliding movements during familiarization. What happened next differed across the two between-subject conditions. In the Sliding- Sliding condition, once the box reached the intermediate target, the black line was stretched out and the box slid to either the near or far target location (Figure 8B). In the Sliding-Jumping condition the black line disappeared (Figure 8C). Then, the box disappeared and reappeared in either the near or far green target location. These changes in the line were introduced in order to highlight the differences in the second occluded movement. The disappearance or stretching of it before the onset of the second movement was used to inform participants about the upcoming disappearance-reappearance or sliding of the box towards the targets, respectively.

During trials, participants in both conditions saw the first sliding movement towards the grey target on the left, and saw either the disappearance of the line or its stretching before the occluder covered the entire layout. Note that the first movement participants saw was identical in both conditions; what differed during trials is the implied (i.e., occluded) second movement only.

3.4.2 Results

Consistency scores. We conducted a 2x3 ANOVA on Consistency scores with second movement type as between subject factor and degrees of exaggeration as a within subject factor. The ANOVA yielded a significant main effect of degrees of exaggeration F(2,94) = 35.9, $p < .001$, $\eta^2 = .25$. The main effect of second movement type did not reach significance F(1,47) = 1.34, $p = .25$, $\eta^2 = .016$, nor did the interaction between these two factors F(2,94) = .18, $p = .81$, $\eta^2 = .002$. (Figure 9).

Mapping scores. We conducted a 2x3 ANOVA with Mapping scores as dependent variable, second movement type as between-subject factor and degrees of exaggeration as within-subject factor. The ANOVA yielded main effects of second movement type ($F(1,47) = 8.3$, $p = .006$, $\eta^2 = .11$), degrees of exaggeration ($F(2,94) = 6.6$, $p = .002$, $\eta^2 = .03$) and an interaction between these factors ($F(2,94) = 3.5$, $p = .03$, $\eta^2 = 02$). Within the Sliding-Jumping condition, none of the pairwise comparisons across degrees of exaggeration using Bonferroni-corrected t-tests yielded a significant difference (all $t(94) > -1.1$, $p > .86$, $d < .09$), whereas in the Sliding-Sliding condition both the comparison between Normal and Exaggerated ($t(94) = -3.4$, $p = .003$, $d = .87$) and between Normal and Very Exaggerated ($t(94) = -4.1$, $p < .001$, $d = .88$) differed significantly (Figure 10).

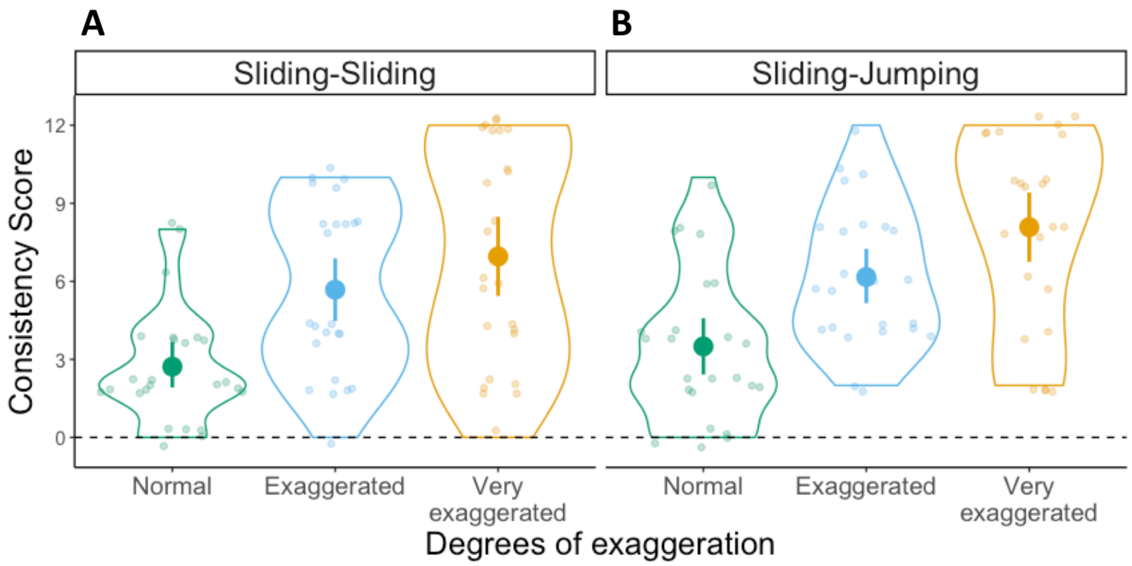

Figure 9. Consistency scores in Experiment 3.

Distribution of Consistency scores in the (**A**) Sliding-Sliding and (**B**) Sliding-Jumping conditions. Violin plots represent the overall distribution of Consistency scores for each degree of exaggeration. The dashed horizontal line indicates a hypothetical value for random mapping (i.e., no consistency).

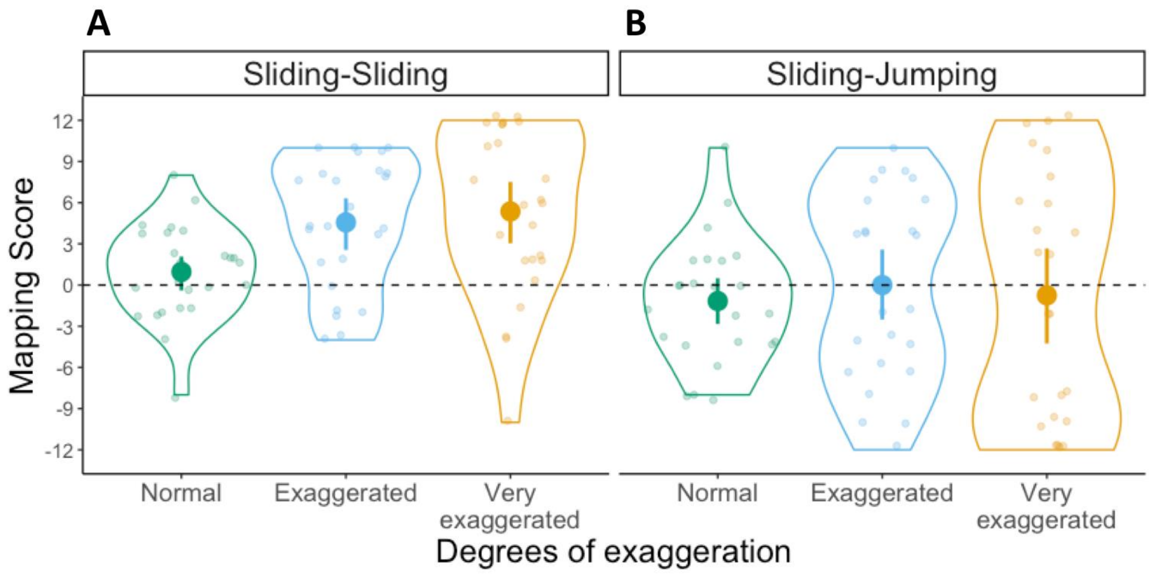

Figure 10. Mapping scores in Experiment 3.

Distribution of Mapping scores in the (**A**) Sliding-Sliding and (**B**) Sliding-Jumping conditions, for the three degrees of exaggeration. Violin plots represent the overall distribution of Mapping scores for each degree of exaggeration. The dashed horizontal line indicates random mapping direction.

3.4.3 Discussion

In Experiment 3 we asked whether highlighting the connection of the second movement to the distal goal would affect how participants simulate a distal goal, and whether such change would also depend on whether the second movement contains velocity information or not. By doing this, we created a situation in which participants would be more likely to integrate the second movement into their simulations (i.e., movement-to-movement simulations). The findings of Experiment 3 suggest that participants do integrate information coming from the second movement component when simulating a distal goal. When presented with continuous movements containing velocity information which are then followed by another continuous movement containing velocity information (i.e., Sliding-Sliding), participants were more likely to map the movements to the target locations in a motor-iconic fashion. But when the same continuous movements were followed by a second discrete movement that did not contain velocity information (i.e., Sliding-Jumping), participants did not show a clear preference towards any mapping direction, and thus failed collectively to simulate the distal goal. Thus, our findings support a more complex conclusion regarding the simulations that observers use to predict a distal goal. In some (and maybe most) circumstances observers can simply rely on movements containing modulations in velocity to run direct movement-to-goal simulations that do not take into account a potential second movement. In others, however, the second movement, particularly when it contains informative kinematic features like velocity, can facilitate the process whereby observers link initial movements to their distal goals. In Experiment 3 we showed that one such situation concerns the reversal of the initial movement, which meant that the direction of the first and second movement were opposed, with the first going away and the second towards the target locations. This meant that, while the first movement still contained useful velocity information that could be used to simulate the distal goal, its change in direction led to a weakened relationship with the distal goal. Possibly, other situations in which the relationship between the movements or between the movements and the distal goal is modified might lead to a similar change in the type of simulations that observers rely on.

In sum, these results support the hypothesis that observers infer distal goals taking into account not only the kinematics of the first movement in a sequence, but also the features of an implied (i.e., occluded) second movement.

3.5 General Discussion

Previous research on SMC focused on settings in which observers made online predictions of unfolding movements in order to derive a proximal goal (Pezzulo & Dindo, 2011; Sacheli et al., 2013; Vesper, Schmitz, et al., 2017; Vesper & Richardson, 2014). Here we extend this focus to situations in which they need to simulate a future, more distal goal (Dockendorff et al., 2023). To do so, we first differentiated proximal and distal goals with respect to the number of motor acts used to achieve them: proximal goals are achieved by a single motor act, such as reaching for an object, while distal goals are achieved by means of at least two sequential motor acts, each with their own proximal (sub-)goal, like reaching for an object to then throw it (Marteniuk et al., 1987; Gentilucci et al., 1997). Following this distinction, we created a task in which we presented participants with the first component of a partially occluded two-step action sequence, and asked them to predict the distal goal of the sequence on the basis of communicative modulations present in the first component. Thus, our task differed from previous studies in SMC in that, instead of presenting one-step actions whose goals needed to be predicted as the movements unfold (see Pezzulo et al., 2019 for a review), we showed participants a full initial movement leading to an intermediate (sub-)goal, and asked them to simulate the distal goal of the entire sequence.

Using this paradigm, we were interested in studying the conditions under which observers can derive information about a distal goal by using two types of simulation: one in which observers use the kinematic information presented in an initial movement to simulate the distal goal directly, which we referred to as “movement-to-goal” simulation, and another in which they use that same kinematic information to simulate an upcoming movement, subsequently leading to a simulation of a distal goal and which we referred to as “movement-to-movement” simulation.

To do so, we first looked at whether the presence or absence of continuous velocity information in the first movement of the sequence would enable observers to simulate a distal goal. Participants in Experiment 1 were presented with either a continuous sliding movement that contained velocity information or a discrete jumping movement that only had information about its total duration. These movements were then followed by a second, occluded sliding movement towards one of two target locations. We found that participants established consistent mappings both when the velocity information was presented and when it was not, and that such mappings became more consistent with higher degrees of exaggeration, in line with previous findings in SMC (Dockendorff et al., 2023; Vesper, Schmitz, et al., 2017). Despite the high consistency of their mappings, participants were only able to simulate a distal goal when the movements they observed contained modulations of velocity, and not when they only contained modulations in total duration, as indicated by the strong preference for motor-iconic mappings in the former but not the latter condition. This preference for motor-iconic mappings is a clear indication that observers are able to identify the underlying lawful relation that connects movements and distance during natural performance (i.e., aiming movements reach higher peak velocity when directed at further targets (Jeannerod, 1984; Gentilucci et al., 1997), and use it to connect a first movement component in a sequence to its distal goal.

However, the findings of Experiment 1 were inconclusive with regard to the type of simulation used by participants to predict the distal goal. Participants may have integrated the second sliding movement into their interpretations, consistent with movement-to-movement simulations, or may have completely disregarded the implied second movement, thus supporting a more direct, movement-to-goal simulation. Thus, to get a better understanding of the type of simulation underlying participant's responses we conducted Experiment 2, where we presented participants with the same movements as in Experiment 1, but these were then followed by a movement that only contained information about duration. Thus, we created a situation in which participants would be less able to use movement-to-movement simulations to predict the distal goal of a movement containing velocity. The findings of Experiment 2 suggest that participants can engage in direct movement-to-goal simulations when observing movements containing velocity, and thus can bypass the second movement in their simulations of the distal goal. However, the results also showed that movement-to-goal simulations were leading to more variability in the way participants mapped the movements onto the target locations. This made us hypothesize that, at least in some circumstances, participants' simulations of a distal goal are sensitive to the presence of a second movement, even if this movement is not directly observed but only implied.

The particular aim of in Experiment 3 was to address this question. In other words, would there be situations in which movement-to-movement simulations also play a role? We hypothesized that one such circumstance would be if the relationship of the first movement with respect to the distal goal is weakened, thus highlighting the role played by the second movement. This weakening of the link between the movements was made possible by reversing the direction of the first movement, which was directed away from the target locations, while the second movement was directed towards them. With this manipulation, we expected participants to integrate the second movement into their simulations, and therefore also expected their responses to be affected by whether this second movement contained velocity information or not. The results of Experiment 3 confirmed this prediction, as participants interpreted the movements presented differently depending on whether the second movement contained velocity information or not.

What does this set of studies tell us about the relationship between proximal and distal goals? One way of describing the relationship between proximal and distal goals, besides the number of motor acts leading to their achievement, is to locate them vertically along an action hierarchy, with overarching (potentially more distal) goals on top, and simpler motor acts (possibly more proximal) at the bottom (Jeannerod, 1994; Csibra, 2007). This way of describing proximal and distal goals is interesting for two reasons. First, it implies that proximal goals are sometimes instrumental for the achievement of more distal goals, as the former are simply the subgoals leading to the achievement of the latter (e.g., the proximal goal of picking up an apple is a subgoal leading to the more distal goal of eating it). Second, and maybe more relevant for our purposes, this hierarchical organization suggests that proximal goals located at different levels of the action hierarchy can be used by observers to simulate distal goals which might also vary along the same hierarchy. For example, the initial stages of an aiming movement towards a ball can then be used not only to predict whether the agent will then throw it into a large box (Marteniuk et al., 1987) but more generally, whether the agent is doing so because she has a goal located high up in the action hierarchy, like tidying up a room or simply practicing her aim. In the context of SMC, where people exaggerate their movements to convey anticipatory information about their action goals and thus facilitate coordination, one would expect observers to be similarly able to derive information about such overarching, higher goals, especially when these are relevant to the achievement of a joint goal. For instance, observers are sensitive to the kinematics of their co-actor's instrumental movements while playing a speeded game, but even more so depending on whether the game is framed by the experimenter as a cooperative game, rather than a competitive one (Lewkowicz et al., 2013), suggesting that higher, and in this case pro-social, goals can have strong top-down effects on people's sensitivity to others' actions (see also Manera, Becchio, Cavallo, et al., 2011; Sartori et al., 2011). Within the context of SMC, observers have been shown to benefit from communicative modulations not only to predict simple action goals, but also to infer whether an actor is performing a given action sequence with the arguably more complex social goal of demonstrating the sequence to a naïve observer (i.e., teaching), or to coordinate with someone (McEllin, Sebanz, et al., 2018). Altogether, these studies suggest that people can derive useful information from observing other people's movements, and from these observations infer more than mere action goals that vary with respect to the number of motor acts (i.e., proximal and distal), but that can also be located at different levels of an action hierarchy, starting from simple motor actions all the way up to complex social intentions.

An open question for future studies is whether communicative modulations of instrumental actions can, on top of facilitating the prediction and identification of more or less complex goals, trigger other types of inference about the observed action. For example, recent computational models of communicative demonstrations have shown that participants can learn the hidden reward structure of grid-like environments when observing movements that deviate from the most efficient trajectory (e.g., by visiting multiple tiles within a trial). Crucially, participants in this task increased their accuracy when told that the agent producing the movements knew that naïve observers would then watch and learn from them (Ho et al., 2021). In line with findings from SMC and action observation, this indicates that observers interpret movements differently if they know that these were produced not only with an instrumental goal, but with a further communicative goal. Another interesting example of observers going beyond the mere prediction of goals is a recent study by Schmitz et al. (2018b), where they showed that observers can infer the hidden properties of an object (i.e., its weight) by relying on communicative modulations of reaching movements directed towards these objects. Thus, these two studies open up a venue for future research on people's capacity to derive other kinds of information when observing communicative modulations, either about the person performing the action (e.g., what the person knows Aboody et al., 2022) or about hidden properties of objects (e.g., their function Hernik & Csibra, 2015).

Extending our own previous research (Dockendorff et al., 2023), we made two proposals about the underlying simulation processes that would enable observers to link communicative modulations in early instrumental movements to their distal goals. These two theoretical possibilities differ primarily in the amount of information needed in order for the simulation to occur. A "movement-to-goal" simulation is made possible when observers link an early movement to its distal goal directly, and thus do not need to take into account the role played by the second movement in the sequence. A "movement-to-movement" simulation, on the other hand, does integrate the second movement in the process of linking the early movement to the distal goal. The process of using the kinematic information present in an initial movement to feed into a simulation of a second movement is in some respects analogous to the process of action simulation described by Prinz & Rapinett (2008), according to which observers use the early kinematic features present in a reaching movement before its occlusion to generate an internal simulation that replicates the kinematic features of the reaching movement and then applies them to the now extrapolated reaching movement occurring behind the occluder (Springer, Parkinson, et al., 2013). This internal simulation, which reuses information from the early stages of the movement, before its occlusion, enables observers to make accurate predictions about the reappearance of the movement after these short episodes of occlusion. As such, movement-to-movement simulations and the action simulation processes described by Prinz and Rappinet (2008) seem to rely on the same underlying principle which enables observers to predict action trajectories and goals by means of internally regenerated movements, i.e. simulations. The main difference between the two processes, however, is that while action simulations are commonly used in the process of extrapolating partially occluded one-step movements, the movement-to-movement simulations we propose here are used to regenerate an entire movement, with its corresponding (distal) goal.

To address the question of how observers can simulate distal goals, we decided to focus on two-step action sequences in which only the first movement was visible to participants. This meant that, unlike previous research on SMC in which proximal movements are exaggerated to facilitate the online prediction of proximal goals, the first movement was temporally and spatially separated from the simulated distal goal. In our previous work (Dockendorff et al., 2023), we discussed the possibility that such separation may have led some participants to change the way in which they interpreted the movements they saw. Specifically, participants in our experiments may have fully disregarded the instrumental aspects of the movement (i.e., delivering the box), while focusing exclusively on its communicative goal (i.e., informing observers about the upcoming delivery location). This could have happened despite the fact that participants were explicitly told that the movements had an instrumental goal at the beginning of the study. If participants fully disregarded the instrumental aspect of the movements, this would imply that they may have taken the movements as purely communicative movements that stand for, represent, or refer to particular target locations, similar to other "purely" communicative movements, like gestures, that are also said to stand for, represent, or refer to particular entities by means of hand and bodily movements. Indeed, functional accounts of gestures have recently put forward the idea that what makes a movement a "gesture" is the fact that these are bodily movements that are stripped away from their more habitual instrumental aspects (e.g., reaching and manipulating objects). In the process of becoming less "instrumental", these movements come to fulfill a different function, essential to most gestures: that of representing or referring to objects (Novack et al., 2016; Novack & Goldin-Meadow, 2017), either for oneself or for others.

This way of understanding the difference between instrumental and communicative actions has interesting implications for our current findings and for SMC more generally. Given that SMC relies on the production and understanding of movements that have both an instrumental and communicative goal, this form of communication can be seen as occupying an intermediate position between fully instrumental and fully communicative movements (Clark, 2005; Pezzulo et al., 2013, 2019; Ho et al., 2021). Consequently, one could argue that the communicative modulations present in SMC might already contain some of the ingredients that enable movements to become "representational", or "referential", thus making them more similar to gestures. For example, in our studies, communicative modulations of an early movement might be seen as referring to the distal goals of the sequence (corresponding, in this case, to specific movement endstates). What makes this relationship between movement and goal one of "reference" is the fact that the movements are separated, and thus "detached", from the goal. From the observer's point of view, this form of detachment might be seen as a first step in the process of interpreting movements as having the capacity to "represent", which subsequently might lead observers to come up with stable mappings between these movements and their goals. Whether and how SMC can provide a standpoint from which to study the relationship between instrumental and communicative actions, and how the former kind of action becomes more like the latter by means of gradually acquiring such "representational" features, are among some of the questions that will require further investigation.

Besides providing evidence that observers can simulate goals that are spatially and temporally removed from the here-and-now (i.e., that are distal), our studies also indicate that these simulations drive participants towards establishing specific mappings between movements and the distance of target locations, where faster movements are more likely to be mapped onto far locations (Gentilucci et al., 1997). In the Discussion of Experiment 1 we mentioned how this motor-iconic mapping can be contrasted with its reversal which, interestingly, was more likely to be present in the conditions in which participants saw movements that only contained information about their total duration (i.e., Jumping). In other words, participants in these conditions were more likely to interpret *longer durations* as directed to *further* target locations. As we pointed out, this mapping was also consistently used by participants in a previous study on SMC, by Vesper et al., (2017), in which pairs of participants were asked to coordinate their actions by performing aiming movements towards one of three target locations. Their results show that Leaders systematically mapped shorter aiming movements to near targets and longer aiming movements to far targets.

Although our study and the one by Vesper and colleagues (2017) differ in many respects, they both point to similarities in the intuitions that people have about the relationship between particular movement parameters (such as duration or velocity) and distances. Thus, in both studies, the predominantly chosen mappings can be considered instances of "motor-iconicity". Whereas most participants in the present study focused on the regularities between movement velocity and movement distance (i.e., faster velocities go with farther distances), the task layout in the Vesper et al. study highlighted the relationship between movement duration and movement distance (i.e., longer durations go with farther distances). What makes this latter mapping motor-iconic is the fact that the duration-distance mapping also originates from a regular relationship found in the performance of aiming movements, where people tend to take longer (in terms of duration) to reach further locations (Fitts & Peterson, 1964). From this point of view, participants in the Jumping-Sliding and Jumping-Jumping conditions who, according to our interpretation, were "reversing" the mapping, might simply rely on a different, but still motor-iconic, relationship connecting the movements to their goals. Whether people are aware of these differences in motor-iconic relations, and whether they can use these to inform others about their goals, are among the questions that should be further investigated.

Chapter 4. Modulating one's early actions to communicate about distal goals

4.1 Introduction

Many joint actions require people to closely monitor and predict what others are doing in order to achieve a joint goal. One important way in which co-actors do this is by observing the kinematics of each other's movements, such as the velocity or amplitude of a reaching movement or the hand aperture during a grasping movement (Becchio et al., 2012; Manera, Becchio, Schouten, et al., 2011; Quesque et al., 2016). From this, they can derive a range of information about each other's actions and goals. Based on this information they can form predictions that allow them to adapt their behavior to what the other is doing or is about to do, typically facilitating interpersonal coordination (Sebanz & Knoblich, 2009).

Moreover, co-actors can and do often support this prediction process further. Specifically, co-actors can facilitate each other's predictions by modulating the kinematics of the movements they execute while performing the joint action, often called "sensorimotor communication" (i.e., SMC, Pezzulo et al., 2013, 2019). For instance, it has been shown that co-actors modulate the duration of their aiming movements to inform others in advance about the relative distance of a target (Vesper, Schmitz, et al., 2017). SMC has so far mostly been studied in contexts in which co-actors provide information to each other about goals that are directly tied to the performed action, like aiming for a target (Vesper & Richardson, 2014) or grasping an object (Sacheli et al., 2013). We call these actions proximal actions, and their goals proximal goals. When predicting the most likely proximal goals of an action, observers rely on the kinematics of the movement while it unfolds in real-time.

Prediction of more complex actions, however, often involves "distal" goals. These are goals that result from two or more sequential action steps, as in reaching-to-throw or reaching-to-give. Previous research in motor control has shown that the initial stages of a two-step action (e.g., the reaching component of a reaching-to-throw action) are affected by distal goals during natural performance (Marteniuk et al., 1987; Gentilucci et al., 1997). This, in turn, allows observers to predict distal goals when presented with the initial stages of the performer's movement (Becchio et al., 2010; Lewkowicz et al., 2013). Based on these findings, in a recent study, we presented participants with the first step of a two-step sliding action and asked them to predict the distal goal (i.e., the movement endpoint) of the whole sequence. As expected, participants were better at predicting the goal when the action they observed was communicatively modulated (Dockendorff et al., 2023, 2024). To this end, most participants relied on the regularities that are known from production of sequential actions (Marteniuk et al., 1987). In particular, participants relied on the regular relationship between movement duration and movement distance, where far target locations are normally reached with faster movements. We refer to these as "motor-iconic" mappings between movement kinematics and distal goals.

In sum, previous findings suggest that people can rely on the observation of an actor's communicative modulations to predict distal goals. In the present set of studies, we shift our focus from the observation to the *production* of these modulations**,** and ask whether and how participants will modulate their instrumental actions to communicate to their co-actors about a distal goal.

To do this, we designed a task involving two-step action sequences. The initial action was directed at an intermediate goal followed by an ensuing action directed at one of two possible distal goals (i.e., two movement endpoints), one of which was the correct goal for that trial. To succeed in choosing the correct distal goal, members of the pair had to work together in two ways. Each member performed one step of the sequence: the "Sender" performed the first step to the intermediate goal and, subsequently, the "Receiver" performed the second step towards the distal goal. Both actions were thus necessary for the achievement of the joint goal. Furthermore, only the Sender knew which of the two distal goals was the correct one. This posed a coordination problem for the pair, since now achieving the correct distal goal depended on the first member figuring out a way to communicate this information to the Receiver. In Experiment 1, we tested whether pairs of participants succeed in solving the coordination problem.

In Experiments 2a and 2b, we further investigated how participants establish a communication system based on movement modulations. Our previous findings suggest that observers prefer to map movement modulations onto distal goals motor-iconically (Dockendorff et al., 2023, 2024). However, this not only leaves open the question of whether producers will also display such a preference, but also whether co-actors will manage to agree on this mapping at all. From this perspective, Experiments 2a and 2b build on previous studies in “experimental semiotics” that look at how people create novel forms of communication from scratch, without the help of language or other pre-established or conventional codes (Galantucci & Garrod, 2011). Many factors contribute to whether and how people can create such communication systems (Nölle & Galantucci, 2022 for a recent review). Here, we will focus on whether Senders and Receivers can alternate their roles throughout the experiment (also known as interchangeability, see Delliponti et al., 2023). Role alternation has been previously argued to be essential in the process of agreeing on the meaning of drawings used to depict various concepts (interactive grounding, Garrod et al., 2007). In Experiments 2a and 2b we extend these findings by looking at how role alternation contributes to establishing a modulation-based communication system (Schmitz et al., 2018b; T. C. Scott-Phillips et al., 2009). In the process, we also investigate how feedback affects the production of communicative signals and the communicative success.

4.2 Experiment 1. Do senders modulate their actions to communicate about distal goals?

The aim of Experiment 1 was to test whether participants communicate information about distal goals by modulating their action kinematics. To do so we asked pairs of participants to slide an animated box in two steps. The first participant ("Sender") had to slide it from an initial location towards an intermediate location. Then, the second participant ("Receiver") had to slide it from the intermediate location towards one of two delivery locations, a "near" or "far" location. Since only the Sender knew which of the two delivery locations was the correct one in a given trial, we expected Senders to modulate their movements to inform the Receivers, who needed to rely on this information to perform the second step and slide the box to the correct delivery location. Furthermore, we predicted that the way participants would try to communicate this information is by relying on "motor-iconicity", i.e., by sliding faster when the correct delivery location was "far", and slower when it was "near".

4.2.1 Methods

Participants. We recruited 10 pairs of participants (Age: M = 27.2; SD = 4.33) via the Research Participant System (i.e., SONA) at Central European University. Six pairs were of mixed gender, three female-only and one male-only. Participants were all fluent English speakers. All experiments were approved by an institutional review board (ref. number 2023-26). The study design and analyses was preregistered on OSF and can be found under the following link: https://osf.io/s8de9

Setup and apparatus. The experimental setup consisted of two large computer monitors (Dell P2416D, 23.8", 60 Hz) placed back-to-back on a table. Both monitors were connected to a 13-inch MacBook Pro laptop (2017, 60 Hz). The displays of both monitors mirrored the laptop's screen, which ran a full-screen Psychopy script (see Animations below).

Two Apple Magic Trackpads (2nd generation) were placed on the table in front of each monitor and connected via Bluetooth to the laptop. An Apple Wireless Keyboard (3rd generation) was placed in front of one of the monitors, also connected via Bluetooth to the laptop.

Finally, a large cardboard of 65 x 55 cm was placed on the table between both monitors. This prevented participants, who were seated in front of each monitor and thus facing each other, from seeing each other during the experiment.

Animations: Animations were created using PsychoPy 3 (v 3.1.5; J. W. Peirce, 2007) and consisted of four black-bordered hexagonal shapes displayed at different locations along the vertical center of the screen: an "initial" location on the left, an "intermediate" location in the middle, and the "near" and "far" delivery locations on the right side of the screen. These four locations were connected by a horizontal black line (see Figure 1). A box was displayed at the beginning of each trial within the initial location. The identity of the "correct" delivery location ("Far" or "Near") was written in black letters at the bottom left-side of the screen.

The cursor was replaced by the animated box, which meant that participants were able to move the box by dragging a finger across the Trackpads placed on the table in front of them. The movements of the box were limited to horizontal movements along the black line, similar to sliding movements along a slider. Moreover, the box could only be moved in one direction, from left to right. If participants tried to slide it in the opposite direction the box remained stationary.

Once the box was dragged inside the intermediate location, it disappeared and reappeared after two seconds. Together with the reappearance of the box, the border of the intermediate location changed its color from black to green, indicating that the box was ready to be moved to one of the delivery locations (Figure 1B). When the box entered a delivery location, the borders of the location changed from black to green.

Procedure. Participants were randomly seated on opposite sides of the table, each facing one of the monitors. Depending on which side of the table they sat on, they had the roles of “Sender” or “Receiver”. However, neither the Experimenter nor the instructions used these terms (as this would have revealed the purpose of the study to participants), and instead referred to them as “P1” and “P2”. The large cardboard placed between the two monitors prevented participants from seeing each other. Once seated, they were instructed to carefully read the instructions printed on a paper. The instructions asked them to imagine that they were both working for a delivery company and that their task was to work together to deliver boxes to one of two delivery locations. It also included detailed information about their task during the first half of the experiment. Once participants had read the instructions, the Experimenter entered the room, repeated the instructions verbally, and then demonstrated the different aspects of the task. Participants completed 10 training trials during which the Experimenter made sure they understood all the details of the procedure.

Shared knowledge condition: After being familiarized with the task, participants were instructed to work together to deliver the box to the correct delivery location without talking to each other. During this stage, both participants were able to see the correct delivery location written on the lower left side of their screens. This meant that Receivers (who completed the 2nd step of the sequence) always knew in advance whether to move to the Near or Far target location. Participants completed 50 trials.

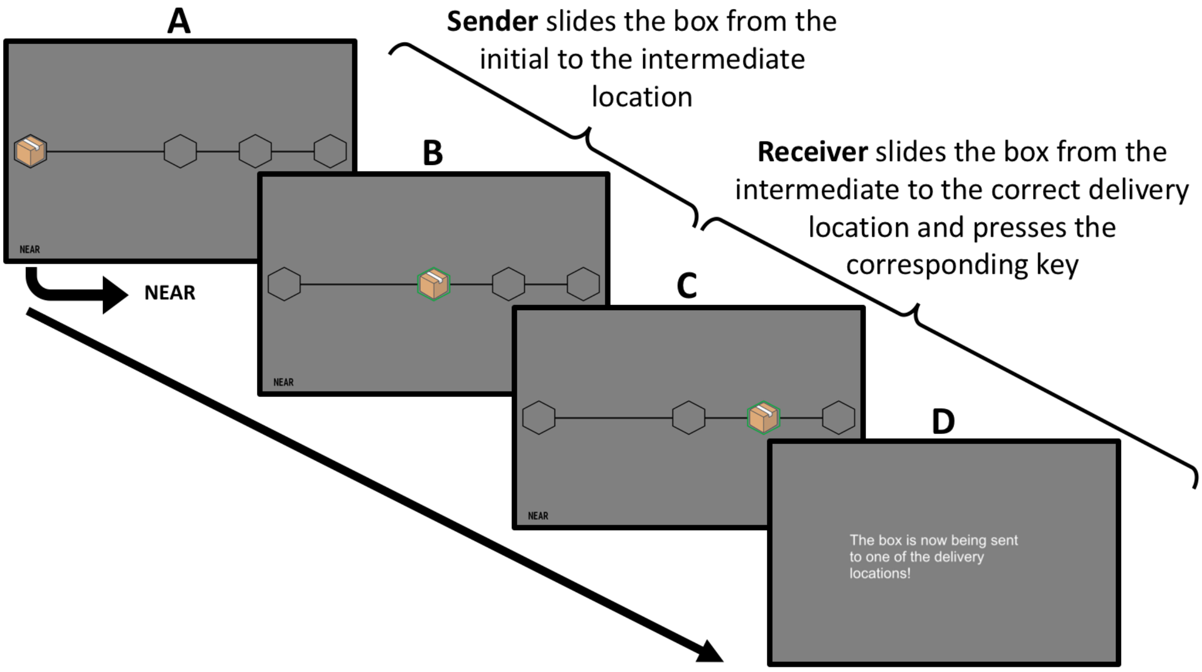

Figure 1. Procedure in Experiment 1.

(**A**) At the beginning of each trial, the box is displayed within the initial location, on the left side of the screen. (**B**) The Sender slides the box from the initial target and places it within the intermediate location. (**C**) Then, the Receiver slides the box from the intermediate location towards one of the two delivery locations (in this case, the near one, as displayed in the bottom left corner of the screen). (**D**) Finally, the Receiver presses one of two keys to confirm the delivery. During the Shared knowledge condition, the Receiver was able to see the correct delivery location written on the lower left side of the screen. This same area was covered during the Partial knowledge condition.

Partial knowledge condition: After completing the Shared knowledge condition, the Experimenter reentered the room and glued a rectangular piece of black cardboard on the lower left corner of the Receiver's screen, thus covering the area in which the correct delivery location appeared. The Experimenter then made sure that both participants were aware that only the Sender (who completed the 1st step of the sequence) was able to see the correct delivery location on his/her screen. The Experimenter told participants that their task was the same as before (i.e., “to work together to deliver as many boxes as possible to the correct delivery location" while not talking to each other). Participants completed 100 trials in the Partial knowledge condition.

In both conditions, the delivery of the box proceeded as follows: Senders were instructed to use their Trackpad to slide the box from the initial location to the intermediate location (Figure 1A). Then, Receivers were instructed to use their Trackpad to slide the box from the intermediate location to the correct delivery location (Near or Far), whose identity either both participants (Shared knowledge condition) or only Senders (Partial knowledge condition) could see written on the lower left side of their screens (Figure 1).

Once Senders placed the box within the intermediate location (Figure 1B), Receivers were instructed to slide it towards the correct delivery location (Figure 1C). To finalize the delivery of the box, Receivers had to press the corresponding key on the keyboard located in front of them ("n" for Near or "f" for Far).

When Receivers had placed the box within one of the delivery locations and pressed one of the keys, both Senders and Receivers saw a text with the message: "The box is now being sent to one of the delivery locations!" (Figure 1D).

At no point did the Experimenter tell participants that they were supposed to communicate in any way. They received no feedback about their accuracy (i.e., whether the box had been delivered to the correct delivery location), but only about whether they had delivered the box. Once participants completed the Partial knowledge condition, they were asked to fill out a short questionnaire about their experience with the task.

Data preparation: We recorded the identity of the correct delivery location, the key pressed by Receivers ("n" or "f" key), and Sender's movements (i.e. spatial location and corresponding time point) with a sampling rate of 60 Hz. To assess the "joint matching accuracy" of the pair during the Partial knowledge condition, we categorized trials in which

Receivers delivered the box to the correct delivery location as correct; the others as incorrect. For convenience, we report joint matching accuracy as percentages.

All the analyses were conducted on RStudio (R Core Team, 2013). From the movement time series, we computed the continuous velocity for all Senders in both the Shared knowledge and Partial knowledge conditions (see Figure 2A). The velocity values were then filtered using a two-directional Butterworth low pass filter with a cut-off frequency of 6 Hz. Based on the entire vector of continuous velocities, we computed, for each trial, the average velocity (see Figure 2B), the peak (i.e., maximum) velocity, and the total duration of Senders' movements.

For each of these three movement parameters we calculated Signal-to-Noise Ratios (SNRs). SNRs capture, in a single metric, the strength and clarity of a signal by looking at the relationship between movement modulations for a given kinematic parameter (i.e., the signal) and their variability (i.e., the noise). Thus, a SNR higher than 1 indicates that the signal is higher than the noise (Vesper, Schmitz, et al., 2017; Vesper, Schmitz, Knoblich, et al., 2016). This is calculated by taking the absolute mean (*M*) difference of each kinematic parameter separately (*kin*: average velocity, peak velocity and duration) for both delivery locations (Far and Near), and dividing it by the mean of their corresponding standard deviations (*SD*), as indicated by the following formula:

$$\mathrm{SNR}_{kin} = \frac{|M\,\mathrm{Far} - M\,\mathrm{Near}|}{M\,(SD\,\mathrm{Far},\, SD\,\mathrm{Near})}$$

We calculated SNRs for each Sender and for each kinematic parameter separately, in both the Shared knowledge and the Partial knowledge conditions. We then selected, among the three SNRs that we calculated for each Sender, the one that reached the highest value during the Partial knowledge condition i.e., the one with the strongest/clearest signal.

Finally, we computed Mapping directions for each Sender in the Partial knowledge condition. We did this by taking the *signed* SNRs, using the following formula:

$$\text{Mapping direction} = \frac{M\,\text{Far - }M\,\text{Near}}{M\,(SD\,\text{Far, }SD\,\text{Near})}$$

The logic of the Mapping direction is the following: if Senders are sliding the box *faster* to when the correct delivery location is *far*, and *slower* when it is *near*, then Mapping directions are positive. Conversely, if Senders are sliding the box *slower* for f*ar*, and *faster* for *near* locations, then it is negative. In other words, if Mapping directions on average are positive, then this indicates that Senders are mapping motor-iconically. Note that for mean durations, the order of the values in the numerator was flipped (i.e., M Near - M Far) in order to obtain a positive Mapping for motor-iconic mappings, and a negative one for reversed mappings.

4.2.2 Results

Figure 2 provides example data of one pair's velocity profiles to demonstrate the analyses we conducted for the whole participant sample. The upper panel (A) shows the velocity profiles of both Senders and Receivers. The velocity profiles are colored depending on the correct delivery location in each trial: blue for Near and red for Far. The left panel shows participant’s movements during the Shared knowledge condition, the right one during the Partial knowledge condition. Descriptively, we can see that in the Shared knowledge condition the Sender does not move differently depending on the correct delivery location (Figures 2A and 2B, left panels), as this information can still be seen by the Receiver. In the Partial knowledge condition, however, we see that Senders clearly modulate the velocity of their movements (Figure 2A and 2B, right panels). In this case, the Sender systematically modulated his/her movement by speeding up when the correct delivery location was Far, and slowing down when it was Near (Figure 2C left panel).

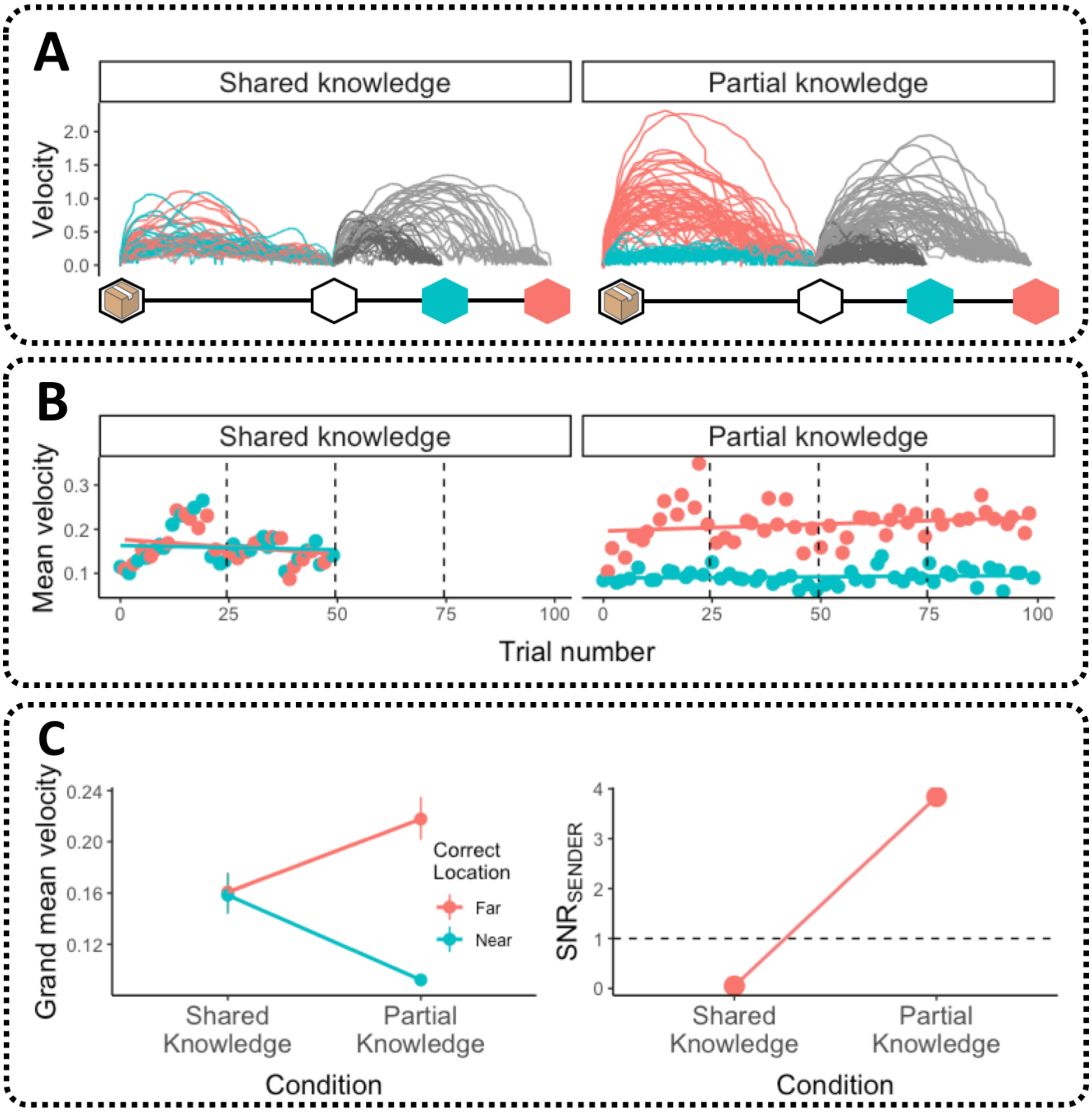

Figure 2. Example data and analysis for one pair.

(**A**) Velocity profiles in the Shared knowledge (left panel) and Partial knowledge (right panel) conditions. Color lines represent the movements performed by Senders within the pair. Grey ones by Receivers within the pair. Red lines correspond to trials in which the correct delivery location was "Far", blue lines to those in which it was "Near". (**B**) Average velocities per trial taken from the same pair in the Shared knowledge (left panel; 50 trials) and Partial knowledge (right panel; 100 trials) conditions. (**C**) Left panel: Grand mean velocity for the same pair, across all trials of the Shared knowledge and Partial knowledge conditions. Right panel: Mean SNR in the Shared knowledge and the Partial knowledge conditions. The black dotted line indicates SNR = 1 (i.e., signal = noise).

Modulation of movement parameters. To get a clearer idea of how Senders modulate their movements, we compared the SNR for each Sender in the Shared knowledge condition to their corresponding SNR in the Partial knowledge condition (Figure 3A). This comparison was significant ($t(9)= -3.69$, $p = 0.004$, $d = 1.63$), indicating that Senders were modulating their movements during the Partial knowledge condition.

Joint matching accuracy. While most Senders seemed to be modulating their movements in the Partial knowledge condition, we found differences in the strength of their modulations, i.e., in their SNRs (see Figure 3A). To see whether these differences in SNR are linked to changes in Receivers' understanding, thereby leading to changes in matching accuracy, we correlated the joint matching accuracy of each pair with the respective Senders' SNR in the Partial knowledge condition (Figure 3B). The correlation was significant ($r(8) = 0.651$, $p = 0.041$), and confirmed previous results in SMC (Vesper, Schmitz, et al., 2017) that point to a relationship between the strength of movement modulations by Senders and the understanding of these modulations by Receivers.

Mapping direction. Now that we found evidence that Senders modulate their movements, we were interested in looking more closely at whether their movement modulations are consistent with "motor-iconicity" i.e., with the differences in movement kinematics present in natural performance. Our results indicate a clear preference for motor-iconicity ($t(9) = 4.13$, $p = 0.001$, $d = 1.31$; one tailed; see Figure 4).

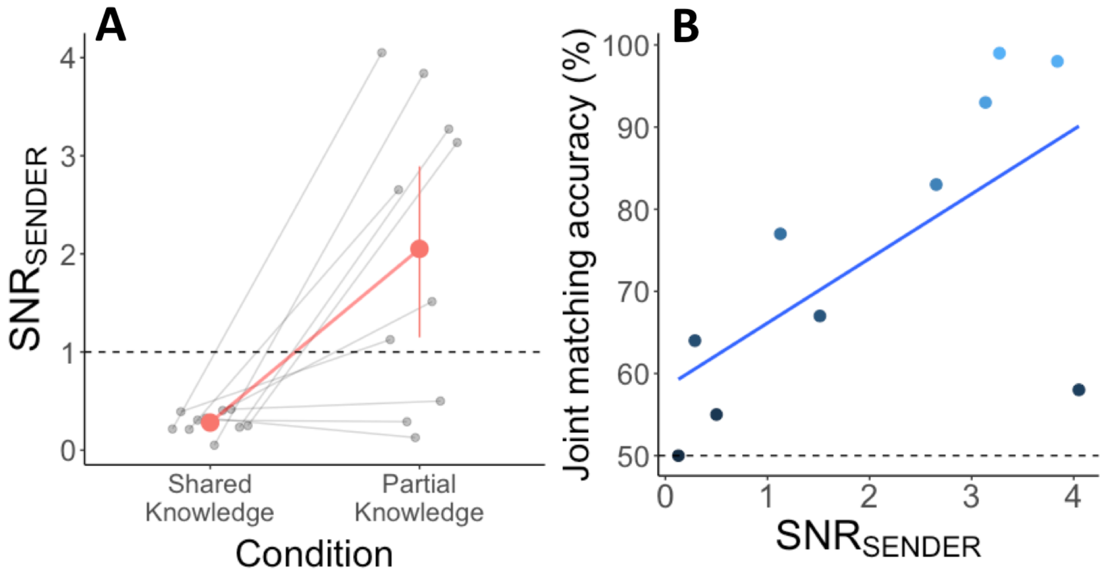

Figure 3. Sender’s modulations and pair’s joint matching accuracy in Experiment 1. (**A**) Movement modulations by Senders in the Shared knowledge and Partial knowledge conditions of Experiment 1. Each grey dot represents an individual Sender, connected by grey lines across the two conditions. Red dots represent mean SNRs for each condition. SNR = 1 is indicated by a black dotted horizontal line (**B**) Correlation of SNRs with the pair’s joint matching accuracy during the Partial knowledge condition of Experiment 1. Each dot is one pair. Lighter dot colors indicate higher joint matching accuracy. Chance behavior at 50 % is indicated by a dotted line.

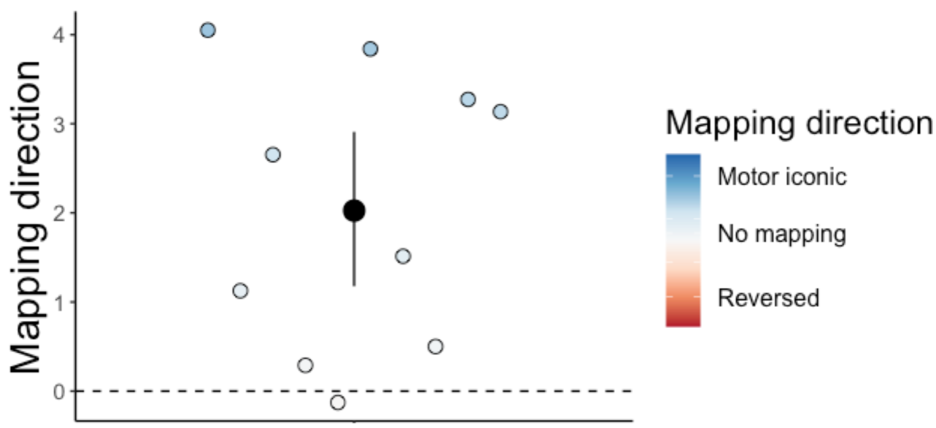

Figure 4. Mapping direction in Experiment 1.

Darker blue colors indicate higher consistency of the motor-iconic mapping, while clearer ones indicate no mapping preference. Dotted black line indicates no mapping preference (i.e., Mapping direction = 0). None of the Senders in Experiment 1 reversed the mapping (in red).

4.2.3 Discussion

The aim of the first experiment was to address the question of whether participants can communicate about distal goals to their co-actors by modulating the kinematics of their instrumental actions. To do so, we created a task in which pairs of participants had to deliver a box in two steps, and where each step of the sequence was allocated to one co-actor. During the Partial knowledge condition, only the co-actor performing the first step of the sequence (i.e., the Sender) had access to the information needed in order for the one performing the second step (i.e., the Receiver) to successfully achieve the joint goal (i.e., delivering the box to the correct delivery location). We expected Senders to provide such information to Receivers, who would then be able to successfully achieve the joint goal.

As expected, most Senders informed Receivers about the correct delivery location by spontaneously modulating the velocity and/or duration of their instrumental actions during the Partial knowledge condition. Receivers, in turn, were able to understand these modulations as conveying information about the locations, leading them to perform their part of the joint action (i.e. sliding the box to the correct delivery location) successfully.

Most Senders relied consistently on motor-iconicity, and in so doing retained the natural differences present in (sliding) movements when they are directed at near and far locations (Gentilucci et al., 1997; Jeannerod, 1984). In other words, they modulated their sliding movements towards the intermediate location in ways that were similar to the way Receivers naturally performed their sliding movements towards the delivery locations (i.e., faster to the far location, slower to the near location, see Receivers' velocity profiles in Figure 2A). These findings corroborate previous findings showing that Senders can use sensorimotor communication to inform Receivers about relevant aspects of the joint task, leading in turn to better coordination. Here, we show that is also the case for distal goals.

Our findings also point to two ways in which pairs can fail to communicate in our task. The first one is when Senders simply do not modulate their sliding movements (see Figure 3B: SNRs close to or below 1 in the Shared knowledge condition). Naturally, this leads to lower joint performance, since Receivers have no signal they can use to identify the correct delivery location. Another failure in communication occurs when Senders modulate their movements communicatively but Receivers fail to understand them (see Figure 3B: SNR > 1 but low joint matching accuracy). In this case, it is not the Sender who fails to communicate, but the Receiver who, for whatever reason, does not comprehend the sender's communication.

These observations raise several questions. One of them is how Senders and Receivers establish a communication system in the first place (Galantucci & Garrod, 2011). Presumably, this process can be facilitated through shared perspectives and common experiences, i.e. if Senders know that Receivers need to be informed about the correct delivery location and Receivers know that Senders can provide such information (Garrod et al., 2007). This leads to the prediction that Senders and Receivers might be able to communicate more efficiently if they are given the possibility to be in both the Sender and Receiver roles in turn. In other words, if they can alternate in their roles. To explore this question, we conducted Experiments 2a and 2b, where we look more closely at how alternating the roles of Sender and Receiver may affect the process of establishing successful communication.

4.3 Experiments 2a and 2b. How do senders and receivers interactively establish a communication system?

With Experiments 2a and 2b, we investigated how Senders establish a communication system based on movement modulations, and whether alternating between the roles of Sender and Receiver affects this process (Galantucci, 2005; Galantucci & Garrod, 2011). To do so, we modified our task such that now participants alternated between the roles of Sender and

Receiver. We did this by grouping trials in four blocks, after each of which Senders became Receivers and vice versa. As in Experiment 1, the Sender was able to see the correct delivery location, and had to slide the box first, to the intermediate location.

In Experiment 2b, participants not only alternated their roles of Sender and Receiver, but they also received further feedback about their correct responses at the end of each trial. While in Experiment 2a the feedback only informed participants about the successful delivery of the box (regardless of whether it was the correct delivery location, see Figure 1D), in Experiment 2b the feedback specified whether the Receiver had delivered the box to the correct or incorrect location. Although the feedback was presented to both participants, we reasoned that it would mainly affect Receivers, since Senders were already able to see if the Receiver was delivering the box to the correct or incorrect location. By investigating whether Receivers would benefit from explicit feedback about their performance, we also aimed at further exploring why some pairs in Experiment 1 failed to communicate.

4.3.1 Methods

Participants. We recruited 12 pairs of participants via SONA for Experiment 2a (3 mixed, 8 women) and 11 on SONA and the Vienna CogSciHub:SPP for Experiment 2b (6 mixed, 5 women only). The conditions for recruiting participants were the same as in Experiment 1 (*M* Age = 25.9, *SD* Age = 2.82). We excluded two pairs in Experiment 2a and one in Experiment 2b because they did not follow the general instructions (e.g., talked during the experiment).

Setup & Animations. The setup was the same one as the one used in Experiment 1. This time, however, the correct delivery location was either displayed on the upper left corner (Blocks 1 and 3), or the lower left corner (Blocks 2 and 4), depending on who had the role of

Sender. Moreover, two Apple Keyboards were connected via Bluetooth, one for each participant (needed when in the Receiver role).

Procedure. The procedure was the same as in Experiment 1 except that trials were grouped into four blocks of 25 trials each (two blocks of 25 trials in the Shared knowledge condition, which had a total of 50 trials; four blocks of 25 trials in the Partial knowledge condition, which had a total of 100 trials). Thus, in the Shared knowledge condition participants alternated their roles as Sender and Receiver once, while in the Partial knowledge condition they did so twice.

As in Experiment 1, after completing the Shared knowledge condition, the Experimenter entered the room and covered the lower left corner of Sender 2's screen and on the upper left corner of Sender 1's screen. Thus, each participant was able to see the correct delivery location in half of the trials (Blocks 1 and 3 for Sender 1, Blocks 2 and 4 for Sender 2, see Figure 5). Between each block of trials, participants were presented with a white cross and a message: "Now P [1 / 2] has to move the box to the middle target!"

In both Experiments 2a and 2b, Senders could observe, in each trial, whether Receivers delivered the box to the correct location, and thus whether the trial was successful. In Experiment 2a, receivers were not informed about whether they delivered the box to the correct location. In Experiment 2b, also the Receiver got informed about the success of the trial. This was done in the form of an explicit message on the screen at the end of each trial. If Receivers delivered the box to the correct delivery location, the feedback consisted of a green cross with the message "Together, you have delivered the box to the correct location!". If the Receiver delivered it to the incorrect location they saw a red cross and the message "Together, you have delivered the box to the incorrect location!". At the end of both experiments participants were asked to fill out a short questionnaire.

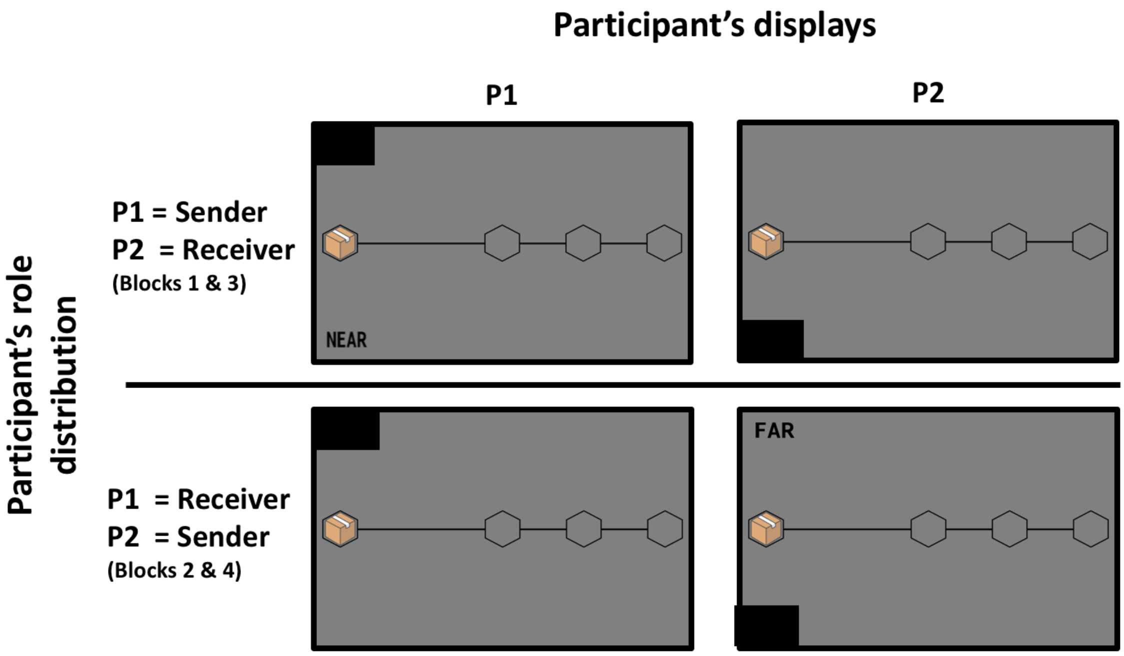

Figure 5. Displays presented to participants (P1 and P2) in Experiments 2a and 2b
Displays presented in the Partial knowledge condition of both experiments. P1 was able to see the correct delivery location displayed on the screen in half of the trials (i.e., Blocks 1 and 3, upper left panel), while Sender 2 was able to see them in the other half (i.e., Block 2 and 4, lower right panel).

4.3.2 Results

Effect of Receiver feedback. We conducted a 2x4 ANOVA of SNR with Receiver feedback as a between-subject factor (Present; Absent) and Block number as a within-subject factor (Figure 6A, grey and dark blue dots). The ANOVA yielded a main effect of Block number ($F(3,54) = 3.61$, $p = .01$, $\eta^2 = 0.07$), but no main effect of Receiver feedback ($F(1,18) = 3.3$, $p = .08$, $\eta^2 = 0.1$) and no interaction ($F(3,54) = 0.32$, $p = .8$, $\eta^2 < 0.05$). We conducted the same analysis on joint matching accuracy and found no significant main effect of Receiver's feedback ($F(1,18) = 1.45$, $p = .245$, $\eta^2 = 0.06$), no significant main effect of Block number ($F(3,54) = 1.42$, $p = .25$, pow = 0.01), and no interaction ($F(3,54) = 1.52$, $p = .22$, $\eta^2 = 0.01$).

Since we did not find significant main effects for Receiver feedback, in what follows we report analyses while collapsing the data of Experiments 2a and 2b (Figure 6, black line). When collapsing Experiments 2a and 2b, and then conducting an ANOVA on both joint matching accuracy and SNR, the significant main effect of Block number on SNR and the non-significant main effect on joint matching accuracy are both preserved ($F(3,57) = 3.74$, $p = .01$, $\eta^2 = 0.06$, and $F(3,57) = 1.38$, $p = .26$, $\eta^2 = 0.01$, respectively) (see Figures 6A and 6B).

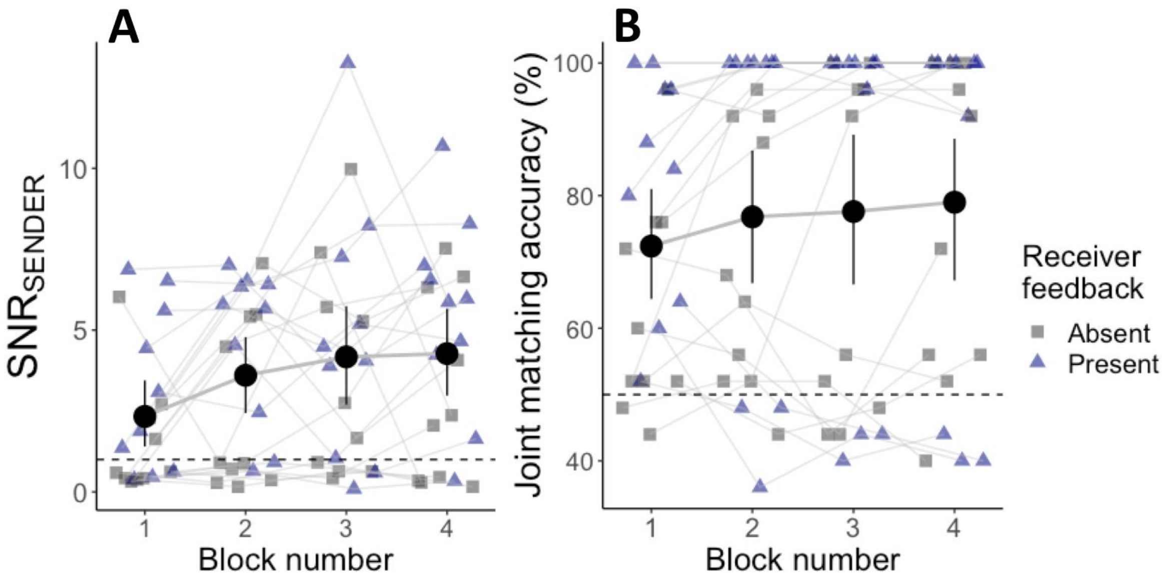

Figure 6. Modulations and joint matching accuracy across the four blocks.
(**A**) SNRs across all four blocks in Experiments 2a and 2b. The dotted black horizontal line indicates SNR = 1 (i.e., noise is as large as the signal). Large black circles represent mean SNR per block. Grey squares represent individual pairs from Experiment 2a, where Receivers did not receive feedback. Blue triangles represent pairs from Experiment 2b, where Receivers received feedback. (**B**). Joint matching accuracy across all four blocks. Chance behavior at 50 % is indicated by a dotted black line.

Effect of early communication. To investigate how early a communication system is established we correlated S1s' SNRs in Block 1 to the pair's joint matching accuracy (Figure 7A). The linear correlation in Block 1 was significant ($r(18) = 0.74$, $p < .001$), confirming our results of Experiment 1.

However, on closer examination of Figure 7A, we noticed that pairs were likely to group around two areas: either close to chance level when the Senders' SNRs were lower than 1 (i.e., when signal = noise), or close to 100% when their SNRs were higher than 1. Because of this, we fitted a logarithmic function to our data in Block 1 (Figure 7A). This function has a steep, positive slope that gradually plateaus when reaching higher values along the y-axis (i.e., the higher the joint matching performance). The logarithmic function revealed to be a much better fit ($r(18) = 0.87$, $p < .001$, $R^2 = 0.63$) than the linear one ($R^2 = 0.53$). There were also high logarithmic correlations in the remaining three blocks (Block 2: $r(18) = 0.79$, $p < .001$, Block 3: $r(18) = 0.76$, $p < .001$; Block 4: $r(18) = 0.80$, $p < .001$; Figure 7B).

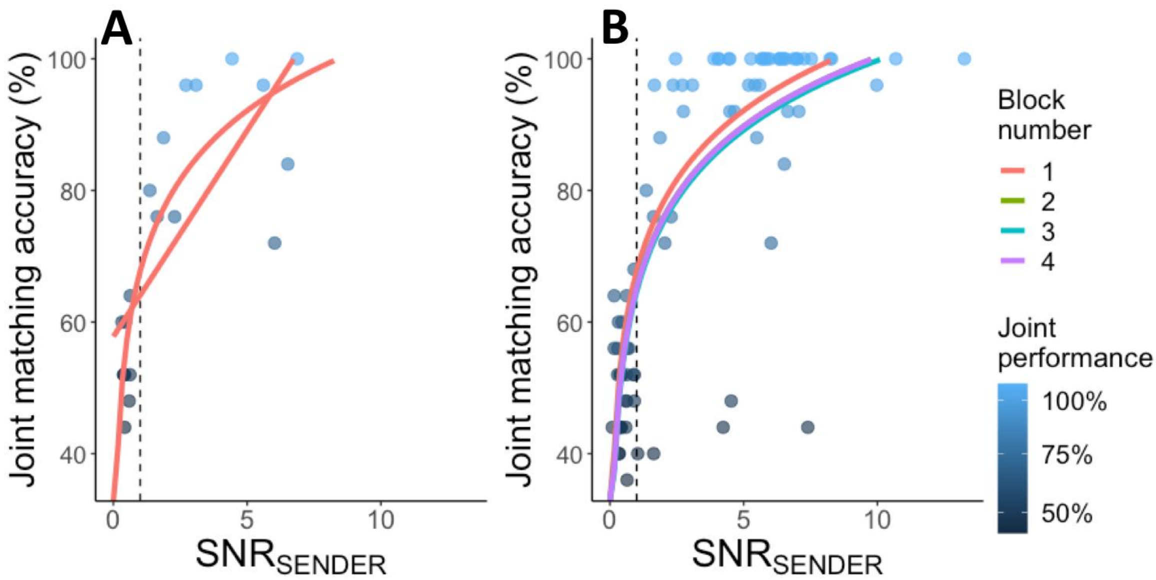

Figure 7. Correlations between movement modulations and joint accuracy.
(**A**) Linear correlation and logarithmic correlations (both in red) in Block 1 between SNRs and joint matching accuracy. (**B**) Logarithmic correlations in all four Blocks (overlapping lines cannot be seen). The vertical black dotted line in both plots indicates SNR = 1.

Given that pairs either seemed to perform well from the start (Block 1, Figure 7A) or had difficulties throughout the experiment (all blocks, Figure 7B), in the remaining analyses, we group pairs according to whether they had early communication attempts or not. More concretely, our criterion for early communication attempt was whether S1s' SNRs were above or below 1 in Block 1, regardless of whether these attempts were understood by Receivers (i.e., not taking into account the joint matching accuracy).

Modulation of movement parameters. We conducted a 2x2x2 ANOVA on Sender's SNR. We included Block stage (i.e., "early" = Blocks 1 and 2; "late" = Blocks 3 and 4) as a within-subject factor and Early communication attempt (i.e., yes = SNR > 1 in Block 1; no = SNR ≤ 1 in Block 1) and Sender (S1; S2) as between-subject factors. This yielded significant main effects of Early communication attempt ($F(1,36) = 40.57$, $p < .001$, $\eta^2 = 0.43$) and of Block stage ($F(1,36) = 9.27$, $p = .004$, $\eta^2 = 0.07$) (Figure 8). None of the other main effects or interactions reached significance (all *p*s > 0.14). The main effect of Block stage indicates that the signals of both Senders are becoming increasingly clearer when going from the early stages (Blocks 1 and 2) to the late stages (Blocks 3 and 4) of the experiment (Figure 8). Furthermore, the main effect of Early communication attempt indicates that, overall, pairs whose Senders communicate early on (in Block 1) produce clearer signals than pairs whose Senders do not communicate at this stage.

Pairs' modulations across blocks. We zoomed in on individual pairs to get a better understanding of how Sender's SNRs change across the experiment. The three individual pairs in Figure 9 below show how SNRs change across the four Blocks as a function of whether they were part of a pair that had an early communicative attempt (Figure 9A and C) or not (Figure 9B). Furthermore, we can also see whether these early attempts led to successfully establishing a communication system later on (Figure 9A) or not (Figure 9C) (See Supplementary Material for all pairs).

Joint matching performance. The ANOVA on joint performance yielded only a significant main effect of Early communication attempt ($F(1,36) = 74.5$ $p < .001$, $\eta^2 = 0.59$) (Figure 10**)**. None of the other main effects or interactions reached significance (all *p*s > 0.17).

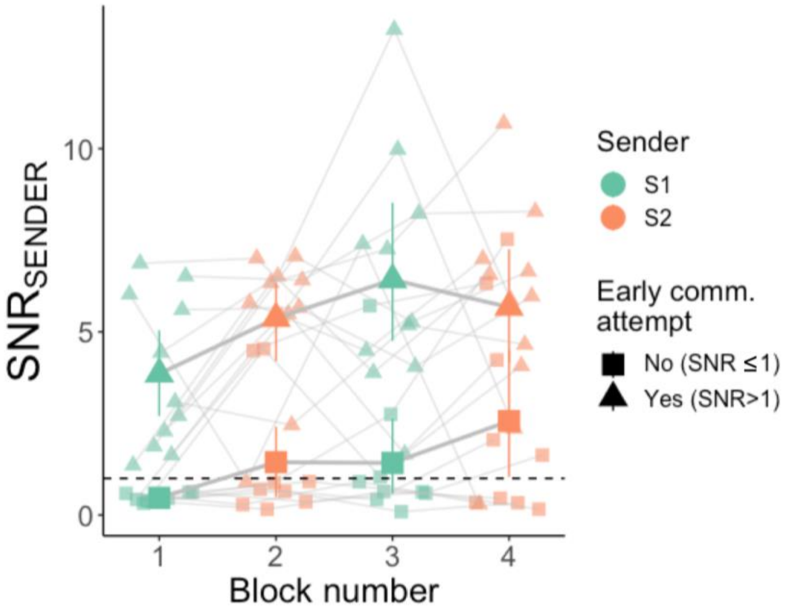

Figure 8. Movement modulations in Experiments 2a and 2b

(**A**) SNRs across all four Blocks depending on whether there is an early communication attempt (triangles) or not (squares). Large figures represent the mean SNR per block. Smaller figures correspond to the mean of each Sender. S1s are colored in green and S2s in orange. Thin grey lines connect Senders within each pair. The dotted horizontal black line shows SNR = 1.

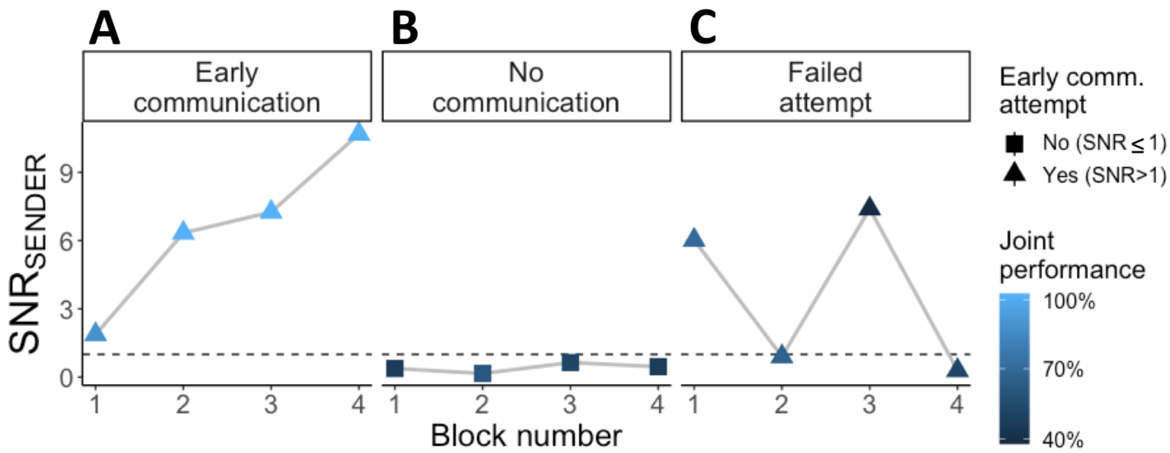

Figure 9. Pairs' movement modulations.

Three dyads illustrating changes in SNR across the experiment. (**A)** S1 communicates in Block 1, thereby leading to higher joint matching accuracy (clearer figures) in the remaining blocks. (**B**) S1 does not communicate in Block 1, leading to no communication and therefore low joint matching performance (darker figures) in all remaining blocks. (**C**) S1 tries to communicate in Block 1 (and Block 3), but the Receiver fails to understand and also fails to communicate back (Blocks 2 and 4).

Mapping direction. To see whether Senders map the movements motor-iconically, we compared the mapping distributions of pairs with and without early communicative attempts to 0 (i.e., no mapping preference) using a one-way t-test. As expected, Senders who did not communicate early on failed, collectively, to display a mapping preference ($t(17) = 1.1$, $p = .14$, $d = 0.25$, one-tailed), whereas those who did communicate were collectively more likely to do so motor-iconically ($t(21) = 3.1$, $p = .002$, $d = 0.66$, one-tailed) (Figure 11).

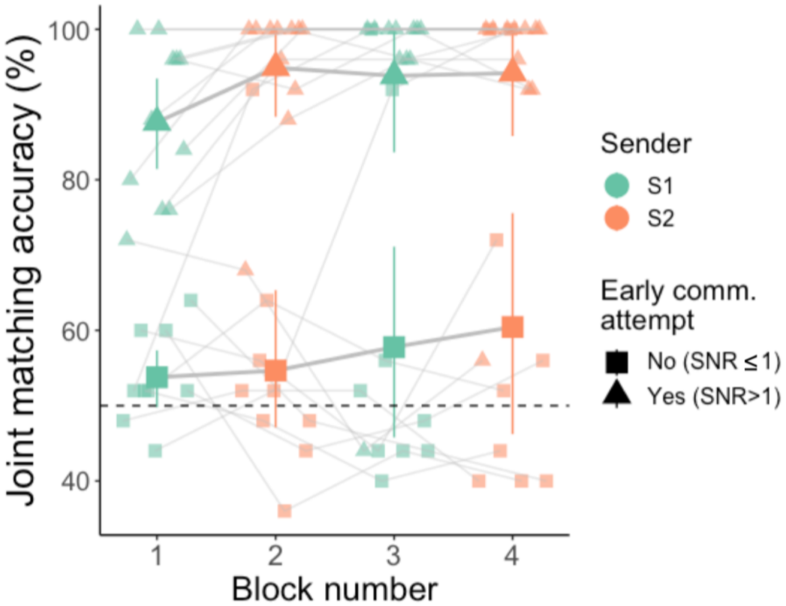

Figure 10. Joint matching accuracy in Experiments 2a and 2b

Joint matching accuracy across all four Blocks depending on whether there is an early communication attempt (triangles) or not (squares). Large figures represent the mean SNR per block. Smaller figures correspond to the mean of each Sender. S1s are colored in green and S2s in orange. Thin grey lines connect Senders within each pair. The dotted horizontal black line shows chance performance per Block (50%).

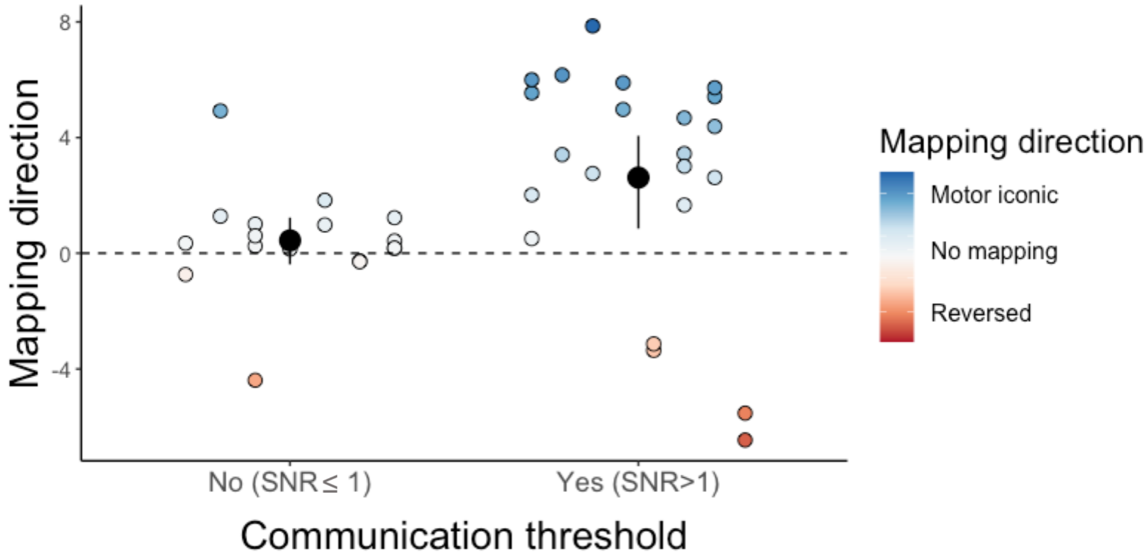

Figure 11. Mapping direction in Experiment 2

Mapping direction based on whether Senders were part of a pair that managed to communicate early on (i.e., Yes (SNR > 1)) or not (i.e., No (SNR ≤ 1)). Darker blue dots correspond to pairs who had a stronger preference for motor-iconic mappings. Darker red dots to those who reversed the mapping. White dots indicate no clear mapping preference. The black dotted horizontal line denotes no mapping preference (i.e., Mapping direction = 0).

4.3.3 Discussion

The aim of Experiments 2a and 2b was to look more closely at the process whereby pairs of participants establish a communication system linking movement modulations and distal goals. Moreover, we were also interested in exploring how, once established, the communication system develops when participants alternate in their roles of Sender and Receiver. Since in Experiment 1 we found that some pairs were failing to communicate, an additional aim in Experiments 2a and 2b was to gain a better understanding of why and how this happens.

We found that the process of establishing a communication system occurs early on in the experiment (i.e., Block 1). Importantly, whether a pair manages to establish a system at this stage has a great impact on whether the pair succeeds at later trials. Specifically, if a Sender in Block 1 discovers a way of communicating (as indicated by a SNR > 1, see Figure 8), the joint matching performance of the pair quickly reaches ceiling performance and remains there for the entire experiment (for an exception, see Figure 9C). But if the Sender fails to do so in Block 1 (as indicated by a SNR ≤ 1, see Figure 8), then in most cases the pair's joint performance remains at chance level in all consecutive Blocks (for two exceptions, see Supplementary Material, pairs 4 and 10). In sum, the findings suggest that pairs succeed at reaching the joint goal if they find a way of communicating early on in the interaction.

We also found that Senders were more likely to increase the clarity of their signals from the first time they acted as Senders (i.e., "early") to the second time they did so (i.e., "late"). This improvement is likely due to two reasons. On the one hand, Senders get to be in the Receiver's role, which gives them first-hand experience with the second sliding movement. Having this motor experience means that, if they are able to notice the kinematic differences between sliding the box to the near or far delivery locations (see Figure 2A, Receiver's movements in grey), they can then "use" these differences later on, when they get to act as Senders. Put simply, being in the Receiver role gives participants the opportunity to experience the "motor-iconic" mapping.

The second likely reason why there might be an increase in signal clarity is that participants get to see how their co-actors fulfill their role as Senders. Thus, if Senders are trying to communicate by modulating their actions in a certain way (e.g., by producing motor-iconic mappings), Receivers can learn from this and adapt their own modulations later on, when they get to be Senders themselves again. This same reason might explain why pairs who fail to establish a communication system early on remain "stuck" in this situation throughout the experiment (see Figure 9B and Supplementary Material). From the Receiver's perspective, seeing a Sender who does *not* communicate is likely to either confirm (if they already think this) or otherwise convince them (if they do not think this) that, actually, no communication is needed.

4.4 General Discussion

The aim of the present study was to investigate whether people can modulate their instrumental actions to communicate about a distal goal to a co-actor. While previous findings in SMC show that people can predict distal goals when merely observing communicative modulations of an early action (e.g., the first step of a two-step action sequence; Dockendorff et al., 2024), here we extend these findings by investigating the interactive side of SMC, with a special focus on the *production* of communicative modulations. Thus, we ask whether actors engaged in a joint action are able to modulate their actions to communicate to a co-actor about a distal goal. Since communicating about the distal goal was, in our task, necessary to successfully achieve a joint goal, we further ask whether and how these modulations lead to better coordination.

To address these questions, we created a task in which pairs of participants slid an animated box in two separate steps towards one of two delivery locations. Each member of the pair had to perform one of the sliding actions: the Sender moved the box from an initial location towards an intermediate location; then the Receiver moved the box from the intermediate location towards one of the two delivery locations, a near one or a far one. During the first half of the experiment (i.e., the Shared knowledge condition), participants had to work together to slide ("deliver") the box to the correct delivery location, whose identity was known to both of them. During the second half of the experiment (i.e., the Partial knowledge condition) the information about the correct delivery location was not available to Receivers any longer. However, pairs were still instructed to "work together" to deliver the box to the correct delivery location. In other words, pairs now faced a coordination problem, since the information relevant to achieving the joint goal was only available to one co-actor (i.e., the Sender) and the success of the joint action depended on this information being shared by both.

In line with a number of studies on SMC, our findings in Experiment 1 show that participants can provide information relevant to achieving a joint goal to their co-actors, and that they do this by modulating the kinematics of their instrumental actions (Pezzulo et al., 2013, 2019; Schmitz et al., 2018b; Sacheli et al., 2013; Vesper, Schmitz, et al., 2017; Vesper & Richardson, 2014; McEllin, Knoblich, et al., 2018; Candidi et al., 2015). Specifically, in our study, Senders modulated the velocity and/or duration of their sliding movements to inform Receivers about the correct delivery location.

By calculating Signal-to-Noise Ratios (SNRs) for each Sender, we were able to measure the clarity of their signals during the Shared and Partial knowledge conditions. We found significant differences in signal clarity between these two conditions. This indicated that during the Shared knowledge condition Senders were, presumably, just "naturally" sliding the box and did not modulate their kinematics as a function of the delivery location (near or far) (see Figure 2A for an illustration). Since at this stage Receivers were able to see the correct delivery location, there was no need for Senders to communicate any information (Figure 2A and 2B, left panel). This changed in the Partial knowledge condition, where most Senders disambiguated their movements depending on whether the correct delivery location was near or far (Figure 2A and 2B, right panel). In other words, most of them understood that they could modulate their instrumental sliding movements to provide information about the delivery locations to Receivers.

Furthermore, we found that clearer signals are linked to better understanding from Receivers, as indicated by higher pair performance (Figures 3B and 7). This confirms previous findings in SMC that suggest that observers benefit from co-actors who are able to produce clearer signals (Vesper, Schmitz, et al., 2017). Finally, we confirmed our own previous findings showing a preference for motor-iconicity when linking communicative modulations of early proximal actions to distal goals (Dockendorff et al., 2023, 2024). Here, we show that this preference is not only present during action observation, but also during production.

In Experiments 2a and 2b, we looked more closely at the process whereby pairs establish a communication system, and how this communication system develops as a result of their interaction (Galantucci, 2005; Galantucci et al., 2012). To this aim, we modified our task such that now each member got the opportunity to act in the role of Sender and Receiver (Garrod et al., 2007). Our findings from Experiment 2 not only confirmed the findings of Experiment 1, specifically about Sender's modulations and Receiver's understanding of these modulations, but they also revealed how pairs establish a communication system using these modulations. This process happens early on and has an important effect on whether the pair keeps communicating, and thereby succeeds in achieving the joint goal later on. In short, the success of the joint action, in the long run, depends on Senders finding a way to establish a communication system early on in the interaction.

What happens when Senders fail to establish a communication system early on? Our findings reveal that when this happens, most Senders keep "naturally" sliding the box until the end of the experiment, without producing any kinematic modulations. This means that they do not try to communicate anything to Receivers. This can be either because they do not realize *that* they can communicate or because they do not know *how* to communicate. Based on our data, we cannot disentangle these two possibilities[2]. However, it is important to note that failures to establish a communication system also occur in other similar settings, especially when people are prevented from using more conventional means to communicate (Galantucci, 2005). Particularly relevant in this respect is a study by Scott-Phillips et al (2009) in which

[2] One of the questions we asked participants at the end of the experiment was: "Did you realize that you were supposed to communicate in order to get more boxes delivered to the correct location?" Among those who did *not* communicate (12 participants, 6 pairs), half replied "Yes". We will refrain from drawing conclusions based on self-reports, since not only they do not reflect participant's actual behavior, but also because we do not know how participants interpreted this specific question (e.g., "I realized, yes, but only now that you are asking me").

pairs of participants had to find a way to communicate to each other while relying on the movements they used to move around a grid. Thus, they had to find a way to communicate *while* moving. Their results were striking: almost half of the pairs (5/12) failed to communicate in the task. In discussing these failures, Scott-Phillips et al (2009) point to the difficulties of having to use the same information channel to move around the grid and to communicate to a partner. These same difficulties might also explain why at least some participants failed in our task.

Taken together, the findings presented here extend previous research in SMC by showing that co-actors can exchange information relevant to achieving joint goals that are not limited to the here-and-now (e.g., proximal goals). Specifically, we show that co-actors can extend their predictions into the future, to include more distal goals around which they can coordinate early on in the interaction. In doing this, Senders rely on mappings that can be readily understood by Receivers, since these mappings preserve something that both Senders and Receivers have experienced in the past (i.e., moving to near and far locations).

Future research could further investigate the scope of these findings. For example, it would be interesting to create longer action sequences and investigate whether there are limitations with how far ahead the distal goal can be relative to the action step that informs about it. How far ahead would participants be able to produce or understand such sensorimotor communication of distal goals?

Another aspect for future research is whether such sensorimotor communication mappings can be generalized to other distal features of an upcoming action. For example, can Senders use the initial step of an action to provide information about the timing at which the Receiver, in a second step, should act? Or more notably: are Senders able to perform action modulations to provide information about an upcoming action that not only the Receiver, but both the Sender and Receiver have to perform *together* (e.g., sliding a box to a delivery location, but jointly)? The findings of the present study provide starting points to address these further questions and help extend our knowledge about sensorimotor communication of distal goals, and more broadly about how communication fulfills a role *within* joint actions and *as a* joint action.

4.5 Supplementary material

4.5.1 Individual pairs in Experiments 2a and 2b

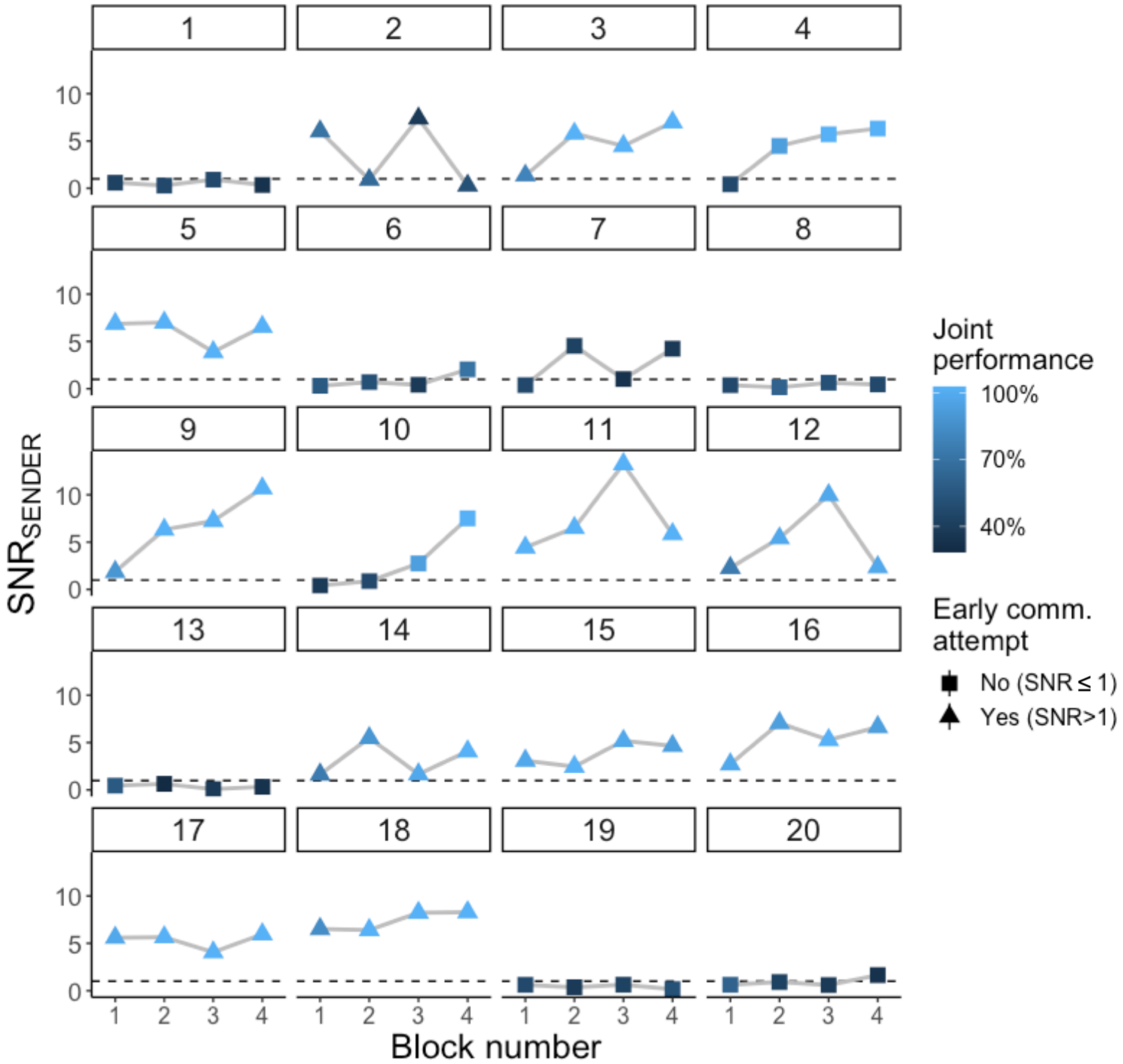

Figure S1. Modulations across the four blocks for each pair in Experiments 2a and 2b
SNR for each individual pair across the four blocks (N = 20). Squares denote the pairs in which there was no early communicative attempt (SNR ≤ 1 in Block 1). Triangles denote the pairs in which there was such a communicative attempt (SNR > 1 in Block 1). The clearer the color of the figure, the higher was the joint matching performance for the pair in that block.

Chapter 5. General Discussion

The general aim of this thesis was to expand on previous research on SMC by asking a series of questions relating to how people use communicative modulations to both predict and communicate about distal goals. These are goals whose achievement goes beyond an immediately observed or performed action (i.e., a proximal action). One example of a distal goal is the final goal of an action sequence made up of two (or more) proximal actions, like reaching for an object followed by passing it to another person (this latter being the distal goal, see Marteniuk et al., 1987).

In Chapters 2 and 3 I presented a set of empirical studies designed to investigate people's *observation* of communicative modulations. First, in Chapter 2, I asked *whether* people can interpret communicative modulations of proximal actions in terms of a distal goal. Then, in Chapter 3, I turned to the question of *how* people predict a distal goal when presented with modulations in the first step of a two-step action sequence. More specifically, I asked what kind of simulation they rely on to make such predictions. Finally, in Chapter 4, I shifted the focus from action observation to production and asked how joint action partners *produce* communicative modulations to inform each other about a distal goal, thereby establishing a novel communication system based on these modulations.

In the following chapter, I will summarize the main findings of my three empirical chapters and discuss the general theoretical implications of the present work. I will also discuss some venues for future research. Then I conclude with a few take-away messages.

5.1 From proximal to distal goals in sensorimotor communication

In **Chapter 2**, I started by asking whether observers are able to understand an early communicative action (i.e., a "proximal action") in terms of its most likely distal goal. Previous research has shown that distal goals can have reliable and visible effects on the kinematics of early proximal actions. These findings, however, are restricted to instrumental (non-communicative) actions and, therefore, leave open the question of how observers interpret early proximal actions when they are communicatively modulated. Specifically, how do they interpret these modulations when they are directed at a distal goal? And finally, do they interpret them in a manner similar to how they interpret modulations directed at a proximal goal?

. To address these questions, I conducted three online experiments where I presented participants with animations of a sliding box being moved at different velocities towards one of two target locations. The box could either be moved directly towards them ("Proximal goal" condition) or it could do so by achieving a more proximal goal first ("Distal goal" condition). Based purely on differences in the velocity of these movements, participants had to predict the most likely (proximal or distal) goal.

The results indicated that participants were able to do this, as shown by their ability to consistently map the movements onto the target locations. In doing so, most participants relied on what I call "motor-iconic" mappings, that is, on a mapping that preserves natural differences in kinematics that are found when people perform reaching or pointing movements towards near or far target locations (i.e., far targets are reached with higher velocities, Jeannerod, 1984). However, these same findings revealed differences in how participants interpret communicative modulations when these are used to communicate about distal goals. Thus, while participants were consistent in their mappings at the individual level, they were split, at the collective level, into those preferring motor-iconic mappings and those preferring the opposite mappings.

Overall, the findings of Chapter 2 confirm previous findings in SMC showing that observers can predict another person's goals when observing modulations in their instrumental actions (McEllin, Knoblich, et al., 2018; Pezzulo et al., 2013, 2019; Pezzulo & Dindo, 2011; Sacheli et al., 2013; Vesper, Schmitz, et al., 2017), while also expanding the scope of these predictions to distal goals. Moreover, the results of Chapter 2 indicate that observers can make these predictions even when they can only rely on simple one-dimensional sliding movements. In this respect, these movements were different to the movements used in previous studies in SMC in that the only information that observers could use to distinguish between them was temporal in nature (e.g., velocity and duration), and not spatial (e.g., amplitude).

More generally, the findings in Chapter 2 make two new contributions. First, they demonstrate that observers can extend their predictions beyond proximal goals to include an actor's more distal goals (Becchio et al., 2018; Donnarumma et al., 2017). Being able to make such long-term predictions about distal goals is particularly useful in the context of joint actions, where co-actors often have to coordinate their actions in order to produce complex action sequences that extend over time, like when two dancers synchronously arrive at the center of a dance floor (Schmitz et al., 2018a).

Second, distal goals change how observers interpret communicative modulations of proximal actions. One hypothesis for this difference has to do with the fact that distal goals, being both temporally and spatially detached from early proximal actions, induce a different interpretation in observers. This "detachment" is the result of one or more intermediate actions that separate the proximal action from the distal goal. In the studies I presented in Chapter 2, the detachment was made even more explicit to participants, since it involved a sudden "teleportation" of the animated box from the intermediate location towards one of the target locations (see Methods Experiment 1, Chapter 2). As a consequence, from an observer's perspective, the relationship between the movement and the goal became less obvious and may have, in turn, induced a more "arbitrary" interpretation of these actions (Schmitz et al., 2018b). Arbitrary, in this context, simply means that observers choose a mapping early on, once they realize that there are systematic differences in movement kinematics that can be mapped onto the action's distal goals. As a consequence, an arbitrary mapping differs from a motor-iconic one in that only the latter involves taking into account the regular relationship between movements and goals, while the former does not.

The idea that distal goals are detached from early proximal actions, and that this can have an impact on how observers interpret these actions, has some interesting implications. As I argued in Chapter 1, SMC differs from other forms of communication in that co-actors combine within a single movement both the instrumental and communicative aspects of their actions (Pezzulo et al., 2013, 2019). This "dual nature" of SMC makes this form of communication very useful in joint actions, since it gives co-actors the possibility to exchange information about their goals *while* they are acting together towards a joint goal. This makes SMC, in short, an ideal coordination smoother (Vesper et al., 2010).

However, in some circumstances, co-actors might actually prefer to keep the instrumental and communicative aspects of their actions apart. For instance, in a study by Vesper and colleagues (2017), "Leaders", who had privileged access to the information needed to achieve the joint goal of aiming towards the same location with their "Followers", increased the duration of their aiming movements to communicate the target location to the Followers. By doing this, Leaders combined the instrumental (i.e., aiming to a target location) and communicative (i.e., informing the Follower) aspects within a single action.

After demonstrating this, the experimenters modified the task such that Leaders now had the option to modify the length of the tone that was played once they hit a target location. This tone lasted for as long as the Leaders kept their fingers on the target (i.e., the dwell time). This change in the task meant that Leaders had now an additional way of informing Followers about the correct target locations, one that did not combine the instrumental and communicative aspects within a single action (i.e., movement duration), but that kept these two apart (i.e., instrumental movement duration on the one hand, communicative dwell time on the other). Leaders, who were given the choice between these two ways of informing Followers, were more likely to choose the one that kept the instrumental and communicative aspects of the action apart. In discussing their results, the authors suggest that this choice helped Leaders to create a more efficient way of communicating to Followers, not only because Leaders had more control over the dwell time than over the duration of their aiming movements (and thus were able to create "clearer" signals), but also because this made it easier for Followers to distinguish between what, in the Leader's movement, counted as a communicative signal and what as a "mere" action (T. C. Scott-Phillips et al., 2009; Royka et al., 2022).

What I would like to suggest here is that when co-actors introduce a spatial and temporal separation between their actions and their goals, as they do when they perform one or more intermediate actions before the (now distal) goal, they give themselves the possibility to separate the communicative from the instrumental aspects across different actions. Thus, an actor can use the initial steps of the action in a manner that is now less constrained by the instrumental aspects of the action. This initial action can become, in the process, more and more schematic (Kita et al., 2017). For example, the actor can use this first step of a sequence to preshape her hand in a manner that indicates the size of an object she wishes to grasp at a later stage. Then, in the second step, she can perform the instrumental action of actually reaching for the object (Kendon, 1991; Goodwin, 2018).

The idea of two actions fulfilling different roles in a sequence fits well with another, more general proposal according to which human social interactions are, fundamentally, structured around sequences. In the particular case I am discussing here, this sequence occurs within a single actor who performs two (or more) actions, each with a different function. But in the case of social interactions, the sequence occurs across two (or more) actors, and can take the form of an action followed by a reaction, or an initiation followed by a response (Kendrick et al., 2020). In other words, organizing actions sequentially may provide a basic infrastructure that makes possible a "division of labor", either within individuals (between communicative and instrumental aspects of individual actions) or across individuals (between a signal and a response) (E. A. Schegloff, 2007).

Going back to the studies presented in Chapter 2, I think that an explanation in terms of a separation of instrumental and communicative aspects could also be given for why some participants adopted a more arbitrary interpretation of the sliding movements when these were directed at a distal goal. Concretely, participants may have chosen a mapping between movements and target locations at the beginning of the experiment because, once they saw that the movements were spatially and temporally detached from their (distal) goals, they interpreted these movements as "purely" communicative, having no "intrinsic" (i.e., motor-iconic) link to the goals. This in turn might explain why participants, although they did not agree collectively on a single mapping, still tried to map the movements consistently to the targets; they knew that there was a signal to be mapped (hence their consistency), but they didn't agree, as a group, on the meaning of the signal.

For now, these observations might raise more questions than answers, and for good reasons. Further research can help in getting a better understanding of the reasons why observers adopt different interpretations of early communicative actions. Is this simply due to the spatio-temporal separation between movements and distal goals? Or is it because participants have a different intuition about the underlying motor-iconic mapping (as I will try to argue below)?

One possibility is that participants, when asked to predict a distal goal on the basis of modulations of a proximal action, may have thought of the proximal action as the first step of an implied two-step action sequence, where the second step is, in this particular case, invisible (because replaced by a “teleportation”). These, and further open questions, were addresses in the studies in Chapter 3.

5.2 Simulations underlying the prediction of distal goals

In **Chapter 3**, I continued my investigation on the observation of communicative modulations. This time, however, instead of asking *whether* observers can interpret these early communicative actions, I turned to the question of *how* they do this. More specifically, I asked what are the conditions under which observers predict distal goals when observing communicative modulations of an early action. In addressing this question, I used a similar paradigm to the one I used in Chapter 2, but this time I focused on two-step action sequences (Lewkowicz et al., 2013). This meant that participants were now familiarized with two separate movement components: an initial movement component from an initial to an intermediate target, followed by an ensuing movement from the intermediate to one of two final targets.

To explain how observers might be able to predict distal goals on the basis of early communicative modulations, I proposed two types of simulation. The first one, movement-to-goal simulation, relies on kinematic information present in the first movement of a sequence to predict, “directly”, a distal goal (Dockendorff et al., 2023). The second one, movement-to-movement simulation, also relies on kinematic information present in the first movement, but this time this information is fed into the simulation of a second movement, which only then leads to a prediction of the distal goal (Prinz & Rapinett, 2008). Consequently, these two simulations differ with respect to the role that the second movement in the sequence plays in predicting a distal goal.

The findings presented in Chapter 3 indicated that observers rely specifically on modulations of movement velocity when predicting distal goals. In other words, only if they were presented with dynamic movement information were they able to predict a distal goal. A movement that only gives away information about its total duration to an observer (i.e., by means of a "jump" indicating only the beginning and end of the movement), even when modulated, is simply not sufficient to run this simulation, and thus cannot be used by observers to predict a distal goal.

When it came to the type of simulation, the findings of two initial experiments indicated that observers were relying on "direct" movement-to-goal simulations, as they were able to predict a distal goal based purely on a single initial movement that contained velocity information. However, the results of a third and final experiment proved this conclusion to be premature. In this experiment, participants were shown a first movement that, although it contained velocity information that could be fed into a simulation of a distal goal, had the peculiarity of moving *away* from the target locations, rather than *towards* them. By reversing the first movement, participants were confronted with a new situation in which, I hypothesized, they would be more likely to integrate the second movement into their simulations (i.e., movement-to-movement simulation). This turned out to be the case, as indicated by the fact that participants formed different mappings depending on the information contained in the second movement.

The findings of Chapter 3 expand those of Chapter 2 by showing that the absence of dynamic information and its replacement with a (implied) movement that only gives away information about its total duration (because participants do not see the full trajectory) prevents participants from predicting a distal goal. Furthermore, the findings suggest that if this information is presented in the early stages of an action sequence, then just observing the movement is sufficient to predict the goal (i.e., "movement-to-goal" simulations). But, if this early movement has, for whatever reason, a "weak" relationship with the distal goal (e.g., because its spatial trajectory moves away from the distal goal), then observers also integrate the second movement into their simulations (i.e., "movement-to-movement" simulations).

The proposed "movement-to-movement" simulation has some interesting parallels with what has been described as a "dynamic simulation" of other people's instrumental actions (Springer, Brandstädter, et al., 2013). The key idea here is that when we observe someone performing an action, we predict the trajectory and possibly the outcome of her actions (i.e., her goals) by running an internal simulation of the observed action in real-time (Graf et al., 2007; Wilson & Knoblich, 2005). Importantly, these real-time simulations can continue even when the observer cannot see the action any longer, as when a reaching movement is occluded halfway through its execution (Springer, Brandstädter, et al., 2013; Springer, Parkinson, et al., 2013; Springer & Prinz, 2010). In this case, the simulation is argued to "substitute" perception by filling in an empty gap left by the sudden occlusion of the action (Prinz & Rapinett, 2008). From this point of view, "movement-to-movement" simulations might rely on similar mechanisms than the ones that enable people to make "dynamic simulations" of other people's movements. Both these processes take, as input, kinematic information coming from a first movement, and then use this information to simulate a second movement.

In discussing these two processes, I have so far focused on how they are both able to integrate specific *kinematic* information (in this case, from a second movement) into a simulation and further prediction of a distal goal. However, an increasing number of studies on action observation shows that observers rely on all sorts of information when predicting an actor's goals, not only on the kinematics of their actions. These include the actor's goals and beliefs (Bach et al., 2014; Bach & Schenke, 2017) objects placed in the vicinity of the actor (Bach et al., 2005, 2014; Schubotz et al., 2014) and environmental constraints (Csibra & Gergely, 2007; Hudson, McDonough, et al., 2018; McDonough et al., 2019).

When it comes to predicting distal goals, a recent study by Thornton & Tamir (2021) showed that observers can use information from actions presented on a video, described in a text or in a movie script to make accurate predictions about the most likely upcoming action.

Rather than based on local information about the agent or her actions, the authors argue that observers can predict distal goals by relying on statistical regularities that connect broad categories of actions in temporal proximity (e.g., stretching, followed by running). All these different sources of information, ranging from an actor's goals to more general knowledge about action categories, raise interesting questions for the study of SMC and its relationship to these other sources of information, particularly when these are used in joint action. For instance, is it easier to predict the actions of someone you hear saying "I am thirsty" while she reaches with a normal movement towards a glass you're holding (Hudson, Bach, et al., 2018), than if you see the same person (silently) modulating her reaching movement towards the glass? How do observers "weigh" and combine these different sources of information?

One final aspect of Chapter 3 that I would like to highlight relates to the mappings that participants produced when they were presented with movements that only contained information about their total duration (i.e., "Jumping"). In such cases, participants were more likely to reverse the motor-iconic mapping, linking *longer* durations onto far targets and *shorter* durations onto near targets. While initially unexpected, these findings made sense in light of other findings in SMC (Vesper, Schmitz, et al., 2017). Specifically, observers might actually not be "reversing" the motor-iconic mapping when presented with a movement that only gives away information about its total duration. Rather, they might be relying on another instance of motor-iconicity, one that highlights movement duration instead of movement velocity.

To illustrate, imagine an actor sliding the box very rapidly towards one of the two locations. If, as an observer, you focus on its *velocity*, you will be more likely to map the movement onto the far target, i.e., motor-iconically. Now imagine that you're presented with the exact same movement, but this time you only hear a short beep when the actor starts moving the box, and then another one when the box reaches the target. Since the movement was very fast, you hear two consecutive beeps separated by a short interval. Here, as an observer, you focus on the *duration* of the movement. If you are anything like the participants who took part in a study by Vesper et al., (2017), you'll be more likely to map *short* durations (like the one I just described) to near targets. What this example illustrates is the fact that, depending on various movement-related aspects (e.g, aiming or sliding) or task-related aspects (e.g., the modality: audio or visual; the presentation method: video-recorded or live) participants might pay attention to different kinematic parameters in the movements (e.g., the duration or the velocity), which could in turn lead to opposite intuitions about how the movements are mapped onto spatial locations (e.g., near or far).

Taken together, the findings of Chapters 2 and 3 converge in showing that observers can interpret communicative modulations present in the early stages of actions (e.g., the first step of a two-step sequence) and link these to a distal goal (Donnarumma et al., 2017). This led observers to create stable mappings between modulated movements and distal goals. From this point of view, the findings reported in both chapters are consistent with other findings in SMC showing that people rely on systematic modulations of movement parameters when creating a novel communication system (McEllin, Knoblich, et al., 2018; Sacheli et al., 2013; Schmitz et al., 2018b; Vesper, Schmitz, et al., 2017). However, these findings also indicate that observers can create such stable mappings while relying exclusively on modulations of temporal parameters (i.e., velocity and duration).

5.3 Modulating one's early actions to communicate about a distal goal

In **Chapter 4**, I presented the results of two experiments designed to expand the focus of the previous two chapters. Here, I turned my attention from the observation towards the *production* of communication modulations in joint action. Like in the two previous two chapters, I asked whether co-actors can communicate about a distal goal by modulating their instrumental actions and, furthermore, how is it that they do it. Finally, I investigated how co-actors are able to establish a stable communication system that maps movements onto distal goals. Since there is no pre-established or conventional mapping that connects movements and distal goals, co-actors need to establish and agree on one if they wish to communicate successfully. In Chapter 4 I aimed at addressing the question of how co-actors manage to do this. More specifically, how is it that they come to agree on a modulation-based communication system?

To address these questions, I presented pairs of participants with a task in which they were asked to collaborate in sliding an animated box in two steps towards a distal goal. Visually, the task was very similar to what I presented to participants in Chapters 2 and 3. The distal goal was to deliver the box to one of two possible delivery locations, a "near" or a "far" location. The first participant (the "Sender") had to slide the box from an initial location and place it inside an intermediate location. Then, the second participant (the "Receiver"[3]) had to slide the box from that location towards the correct delivery location. The identity of the correct delivery location was visible for both participants during the first half of the experiment (the "Shared knowledge" condition) but hidden from the Receiver in the second half of the experiment (the "Partial knowledge" condition). Because of this, Receivers were now not able to deliver the box to the correct location, and thus could not achieve the (joint) distal goal successfully. This posed a "coordination problem" for the pair.

As I expected, most Senders addressed this problem by modulating their sliding movements during the Partial knowledge condition but not in the Shared knowledge condition. Modulating their movements enabled Senders to inform Receivers about the correct delivery location and led to better coordination. Importantly, the clearer the signals produced by Senders, the better they were understood by Receivers. Furthermore, the results of a second

[3] We did not use these labels in the experiment to refer to participants. This would have revealed the purpose of the study. Instead, we simply called them "P1" (Participant 1) and "P2" (Participant 2).

experiment revealed that the process of establishing a successful communication system occurs early on in the interaction, and that this strongly affects whether the pair manages to communicate later on.

These results confirm previous findings in SMC that show that actors can spontaneously modulate their instrumental actions to inform their co-actors about their proximal goals (Pezzulo & Dindo, 2011; Sacheli et al., 2013; Vesper, Schmitz, et al., 2017) or about the property of objects (Schmitz et al., 2018b). Here, I extend these results by showing that co-actors can also modulate their actions to inform others about distal goals. To communicate this information, Senders had to modulate their proximal actions (i.e., the first step of an action sequence) directed at an intermediate target location. However, unlike previous studies in SMC, these modulations were not used by Senders to communicate about their (own) proximal goals. Rather, they were used to communicate about a future, more distal goal which, as it turns out, wasn't even achieved by them (but by the Receiver).

Besides being able to communicate about distal goals, the findings presented in Chapter 4 confirmed the strong preference for motor-iconicity when mapping communicative modulations of early actions to distal goals. This preference is, as I argued in Chapters 2 and 3, strongly present in people's interpretation of these movements during *observation*. The findings of Chapter 4 suggest that this preference also holds for the *production* of these movements.

In Chapter 4 I also investigated the process whereby pairs establish a stable communication system based on movement modulations. In this respect, Chapter 4 builds on previous research on "experimental semiotics" (Galantucci, 2005; Galantucci et al., 2012) that looks at how people manage to find ways of communicating without the help of pre-established conventions or codes. Research in this area indicates that the way in which participants establish a form of communication is strongly affected by whether they are given the opportunity to switch between the roles of director (i.e., sender) and matcher (i.e., receiver) at some point during their interaction (see interchangeability, in Delliponti et al., 2023; Garrod et al., 2007). Based on these previous findings, I asked in Chapter 4 whether role alternation would contribute to how participants establish a communication system.

With respect to the question of how communication is established, the findings revealed that this occurs early on in the interaction. Specifically, if a pair managed to communicate in the early stages of the experiment, and did so successfully (i.e., in ways leading to better coordination), then the pair would then keep doing this for the rest of the experiment. Thus, the success of the pair at coordinating depended, ultimately, on them finding a way to communicate early on.

These results, apart from confirming the function of SMC as a means to facilitate the achievement of joint goals (i.e., a coordination smother), also reveal that the process whereby pairs agree on a shared communication system can be seen as a form of joint action in itself (Chapter 1). This is because agreeing on a modulation-based communication system requires the contribution of both Senders and Receivers. Senders need to "present" a signal by modulating their movements and Receivers need to "accept" the signal by, in this case, completing the joint action (i.e., delivering the box to the correct delivery location). This process is also known as grounding (Clark, 1996; Clark & Brennan, 1991). In Chapter 1, I argued that grounding is not only an essential feature of human communication, but that it also provides a clear illustration of how communication is a form of joint action in which speakers (in the case of spoken language) coordinate their communicative acts to achieve mutual understanding (Wilkes-Gibbs & Clark, 1992). What the findings of Chapter 4 illustrate is that grounding also takes place when co-actors need to understand each other while relying on communicative modulations (see Schmitz, 2017; Schmitz et al., 2018b for similar proposals) demonstrating that this process is, in itself, also a joint action.

The findings of Chapter 4 also revealed that, when pairs fail to establish a communication system early on, then it becomes very hard for them to find a way to communicate later in the experiment. This failure to communicate naturally leads pairs to fail at performing the joint action successfully. In discussing the results of Chapter 4, I argued that cases of failure at communicating are frequently reported in studies of experimental semiotics, particularly when participants cannot exchange information through a specialized communication channel (T. C. Scott-Phillips et al., 2009). The lack of such a channel might also be one of the reasons why some pairs in the studies I presented in Chapter 4 kept just "sliding the box" until the end of the experiment and did not try to communicate.

Besides the lack of a specialized communication channel, there is another aspect of SMC that might also provide some clues about why some pairs failed to communicate. This has to do with the possibility that different joint action contexts might offer co-actors more or less obvious ways and opportunities to modulate their actions communicatively. For example, a joint action context in which co-actors can see each other's upper bodily movements might offer more obvious opportunities for them to modulate their movements than, say, a context in which co-actors can only see each other's cursor movements on a screen. Similarly, the presence of objects that can be reached might also prompt co-actors to realize that they can modulate, for example, their reaching movements towards these objects (Clark, 2003; Schmitz et al., 2018b; Vesper et al., 2021). My point here is that these different aspects of a joint action might lead to differences not only in *how* co-actors use communicative modulations (e.g., via motor-iconicity, symbolically, etc.), but also on *whether* they identify the opportunity to do so. Understanding which aspects of joint actions affect people's sensitivity to these opportunities remains an important question for future research.

Taken together, the findings of Chapter 4 confirm previous studies on SMC that show that participants can provide information relevant to achieving a joint goal to their co-actors, and that they do this by modulating their instrumental actions (Pezzulo et al., 2019; Sacheli et al., 2013; Schmitz et al., 2018b; Vesper, Schmitz, et al., 2017; Vesper & Richardson, 2014; McEllin, Knoblich, et al., 2018). Moreover, the findings provide novel evidence that co-actors can use these modulations to inform a co-actor about a distal goal, i.e., the final goal of a sequence. Finally, I investigated how pairs of participants establish a communication system based on communicative modulations of their actions. In this respect, the findings suggest that successful coordination depends on successful communication, which in turn depends on this latter occurring early on in the interaction.

5.4 Conclusion

How do people use communication to facilitate coordination? This broad question is at the core of this thesis and it provided the background for investigating SMC as a primary example of communication being used for these purposes, i.e., to facilitate coordination. The findings reported here show that people are able to both understand and produce communicative modulations of actions. This gives people the possibility to both predict and inform others about goals that are removed from the here-and-now, as is the case for distal goals. Being able to make this sort of long-term predictions can be of great use when preparing to perform a joint action with another person. But even beyond its role *within* joint actions, the study of SMC can provide researchers with new insights into the processes that underlie communication more broadly, and in so doing, offer new ways of understanding communication *as a* form of joint action.

References

Aboody, R., Huey, H., & Jara-Ettinger, J. (2022). Preschoolers decide who is knowledgeable, who to inform, and who to trust via a causal understanding of how knowledge relates to action. *Cognition*, *228*, 105212. https://doi.org/10.1016/j.cognition.2022.105212

Adkins, T. J., & Lee, T. G. (2021). Reward modulates cortical representations of action. *NeuroImage*, *228*, 117708. https://doi.org/10.1016/j.neuroimage.2020.117708

Ahlheim, C., Stadler, W., & Schubotz, R. I. (2014). Dissociating dynamic probability and predictability in observed actionsâ€"an fMRI study. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00273

Allwood, J. (2002). Bodily Communication Dimensions of Expression and Content. In B. Granström, D. House, & I. Karlsson (Eds.), *Multimodality in Language and Speech Systems* (Vol. 19, pp. 7–26). Springer Netherlands. https://doi.org/10.1007/978-94-017-2367-1_2

Ansuini, C., Cavallo, A., Koul, A., D'Ausilio, A., Taverna, L., & Becchio, C. (2016). Grasping others' movements: Rapid discrimination of object size from observed hand movements. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 918–929. https://doi.org/10.1037/xhp0000169

Ansuini, C., Giosa, L., Turella, L., Altoè, G., & Castiello, U. (2008). An object for an action, the same object for other actions: Effects on hand shaping. *Experimental Brain Research*, *185*(1), 111–119. https://doi.org/10.1007/s00221-007-1136-4

Armbrüster, C., & Spijkers, W. (2006). Movement Planning in Prehension: Do Intended Actions Influence the Initial Reach and Grasp Movement? *Motor Control*, *10*(4), 311–329. https://doi.org/10.1123/mcj.10.4.311

Bach, P., Knoblich, G., Gunter, T. C., Friederici, A. D., & Prinz, W. (2005). Action Comprehension: Deriving Spatial and Functional Relations. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 465–479. https://doi.org/10.1037/0096-1523.31.3.465

Bach, P., Nicholson, T., & Hudson, M. (2014). The affordance-matching hypothesis: How objects guide action understanding and prediction. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00254

Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass*, *11*(7), e12312. https://doi.org/10.1111/spc3.12312

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, *27*(2), 195–225. https://doi.org/10.1016/S0364-0213(02)00118-0

Bavelas, J. B. (2022). *Face-to-face Dialogue: Theory, Research, and Applications*. Oxford University Press.

Becchio, C., Koul, A., Ansuini, C., Bertone, C., & Cavallo, A. (2018). Seeing mental states: An experimental strategy for measuring the observability of other minds. *Physics of Life Reviews*, *24*, 67–80. https://doi.org/10.1016/j.plrev.2017.10.002

Becchio, C., Manera, V., Sartori, L., Cavallo, A., & Castiello, U. (2012). Grasping intentions: From thought experiments to empirical evidence. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00117

Becchio, C., Sartori, L., Bulgheroni, M., & Castiello, U. (2008). The case of Dr. Jekyll and Mr. Hyde: A kinematic study on social intention. *Consciousness and Cognition*, *17*(3), 557–564. https://doi.org/10.1016/j.concog.2007.03.003

Becchio, C., Sartori, L., & Castiello, U. (2010). Toward You: The Social Side of Actions. *Current Directions in Psychological Science*, *19*(3), 183–188. https://doi.org/10.1177/0963721410370131

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, *52*, 46–57. https://doi.org/10.1016/j.wocn.2015.04.004

Branigan, H. P., Pickering, M. J., Stewart, A. J., & McLean, J. F. (2000). Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition*, *28*(8), 1297–1302. https://doi.org/10.3758/bf03211830

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brown, P., & Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage*. Cambridge University Press.

Butterfill, S. A., & Sinigaglia, C. (2014). Intention and Motor Representation in Purposive Action. *Philosophy and Phenomenological Research*, *88*(1), 119–145. https://doi.org/10.1111/j.1933-1592.2012.00604.x

Candidi, M., Curioni, A., Donnarumma, F., Sacheli, L. M., & Pezzulo, G. (2015). Interactional leader–follower sensorimotor communication strategies during repetitive joint actions. *Journal of The Royal Society Interface*, *12*(110), 20150644. https://doi.org/10.1098/rsif.2015.0644

Cattaneo, L., Caruana, F., Jezzini, A., & Rizzolatti, G. (2009). Representation of Goal and Movements without Overt Motor Behavior in the Human Motor Cortex: A Transcranial Magnetic Stimulation Study. *The Journal of Neuroscience*, *29*(36), 11134–11138. https://doi.org/10.1523/JNEUROSCI.2605-09.2009

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific Reports*, *6*(1), 37036. https://doi.org/10.1038/srep37036

Clark, H. H. (1996). *Using Language* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

Clark, H. H. (2003). Pointing and placing. In *Pointing: Where language, culture, and cognition meet* (pp. 243–268). Lawrence Erlbaum Associates Publishers.

Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse Studies*, *7*(4–5), 507–525. https://doi.org/10.1177/1461445605054404

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association. https://doi.org/10.1037/10096-006

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Cohen, R. G., & Rosenbaum, D. A. (2004). Where grasps are made reveals how grasps are planned: Generation and recall of motor plans. *Experimental Brain Research*, *157*(4), 486–495. https://doi.org/10.1007/s00221-004-1862-9

Csibra, G. (2007). *Action mirroring and action interpretation: An alternative account: Vol. XXII* (P. Haggard, Y. Rosetti, & M. Kawato, Eds.; pp. 435–459). Oxford University Press. https://eprints.bbk.ac.uk/id/eprint/29467/

Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*(1), 60–78. https://doi.org/10.1016/j.actpsy.2006.09.007

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447. https://doi.org/10.1016/j.tics.2010.06.009

De Ruiter, J., Mitterer, H., & Enfield, N. (2006). Projecting the End of a Speaker's Turn: A Cognitive Cornerstone of Conversation. *Language*, *82*, 515–535. https://doi.org/10.1353/lan.2006.0130

Delliponti, A., Raia, R., Sanguedolce, G., Gutowski, A., Pleyer, M., Sibierska, M., Placiński, M., Żywiczyński, P., & Wacewicz, S. (2023). Experimental Semiotics: A Systematic Categorization of Experimental Studies on the Bootstrapping of Communication Systems. *Biosemiotics*, *16*(2), 291–310. https://doi.org/10.1007/s12304-023-09534-x

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, *4*(11), 423–431. https://doi.org/10.1016/S1364-6613(00)01537-0

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180. https://doi.org/10.1007/BF00230027

Dingemanse, M., & Enfield, N. J. (2015). Other-initiated repair across languages: Towards a typology of conversational structures. *Open Linguistics*, *1*(1). https://doi.org/10.2478/opli-2014-0007

Dingemanse, M., & Enfield, N. J. (2024). Interactive repair and the foundations of language. *Trends in Cognitive Sciences*, *28*(1), 30–42. https://doi.org/10.1016/j.tics.2023.09.003

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S. C., Manrique, E., Rossi, G., & Enfield, N. J. (2015). Universal Principles in the Repair of Communication Problems. *PLOS ONE*, *10*(9), e0136100. https://doi.org/10.1371/journal.pone.0136100

Dockendorff, M., Schmitz, L., Vesper, C., & Knoblich, G. (2023). Understanding others' distal goals from proximal communicative actions. *PLOS ONE*, *18*(1), e0280265. https://doi.org/10.1371/journal.pone.0280265

Dockendorff, M., Schmitz, L., Vesper, C., & Knoblich, G. (2024). Communicative modulations of early action components support the prediction of distal goals. *PLOS ONE*, *19*(6), e0306072. https://doi.org/10.1371/journal.pone.0306072

Dockendorff, M., Sebanz, N., & Knoblich, G. (2019). Deviations from optimality should be an integral part of a working definition of SMC: Comment on "The body talks: Sensorimotor communication and its brain and kinematic signatures" by Pezzulo et al. *Physics of Life Reviews*, *28*, 22–23. https://doi.org/10.1016/j.plrev.2019.01.010

Donnarumma, F., Dindo, H., & Pezzulo, G. (2017). Sensorimotor Coarticulation in the Execution and Recognition of Intentional Actions. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00237

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130301. https://doi.org/10.1098/rstb.2013.0301

Emmorey, K., Grabowski, T., McCullough, S., Damasio, H., Ponto, L., Hichwa, R., & Bellugi, U. (2004). Motor-iconicity of sign language does not alter the neural systems underlying tool and action naming. *Brain and Language*, *89*(1), 27–37. https://doi.org/10.1016/S0093-934X(03)00309-2

Finnegan, R. (2013). *Communicating: The Multiple Modes of Human Communication* (2nd ed.). Routledge. https://doi.org/10.4324/9781315869872

Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology*, *67*(2), 103–112. https://doi.org/10.1037/h0045689

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal Lobe: From Action Organization to Intention Understanding. *Science*, *308*(5722), 662–667. https://doi.org/10.1126/science.1106138

Foster-Cohen, S. H. (2004). Relevance Theory, Action Theory and second language communication strategies. *Second Language Research*. https://doi.org/10.1191/0267658304sr242oa

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, *8*(1), 113–133.

Galantucci, B. (2005). An Experimental Study of the Emergence of Human Communication Systems. *Cognitive Science*, *29*(5), 737–767. https://doi.org/10.1207/s15516709cog0000_34

Galantucci, B., & Garrod, S. (2011). Experimental Semiotics: A Review. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00011

Galantucci, B., Garrod, S., & Roberts, G. (2012). Experimental Semiotics. *Language and Linguistics Compass*, *6*(8), 477–493. https://doi.org/10.1002/lnc3.351

Galaro, J. K., Celnik, P., & Chib, V. S. (2019). Motor Cortex Excitability Reflects the Subjective Value of Reward and Mediates Its Effects on Incentive-Motivated Performance. *The Journal of Neuroscience*, *39*(7), 1236–1248. https://doi.org/10.1523/JNEUROSCI.1254-18.2018

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218. https://doi.org/10.1016/0010-0277(87)90018-7

Garrod, S., Fay, N., Lee, J., Oberlander, J., & Macleod, T. (2007). Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cognitive Science*, *31*, 961–987. https://doi.org/10.1080/03640210701703659

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11. https://doi.org/10.1016/j.tics.2003.10.016

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Gentilucci, M., Negrotti, A., & Gangitano, M. (1997). Planning an action: *Experimental Brain Research*, *115*(1), 116–128. https://doi.org/10.1007/PL00005671

Georgiou, I., Becchio, C., Glover, S., & Castiello, U. (2007). Different action patterns for cooperative and competitive behaviour. *Cognition*, *102*(3), 415–433. https://doi.org/10.1016/j.cognition.2006.01.008

Gibbs, R. W. (1983). Do people always process the literal meanings of indirect requests? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(3), 524–533. https://doi.org/10.1037/0278-7393.9.3.524

Goebl, W., & Palmer, C. (2009). Synchronization of Timing and Motion Among Performing Musicians. *Music Perception*, *26*(5), 427–438. https://doi.org/10.1525/mp.2009.26.5.427

Goodwin, C. (2018). *Co-Operative Action*. Cambridge University Press.

Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M., & Prinz, W. (2007). Predicting point-light actions in real-time. *NeuroImage*, *36 Suppl 2*, T22-32. https://doi.org/10.1016/j.neuroimage.2007.03.017

Grice, H. P. (1957a). Meaning. *Philosophical Review*, *66*(3), 377–388. https://doi.org/10.2307/2182440

Grice, H. P. (1957b). Meaning. *The Philosophical Review*, *66*(3), 377–388. https://doi.org/10.2307/2182440

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Haggard, P. (1998). Planning of action sequences. *Acta Psychologica*, *99*(2), 201–215. https://doi.org/10.1016/S0001-6918(98)00011-0

Hernik, M., & Csibra, G. (2015). Infants learn enduring functions of novel tools from action demonstrations. *Journal of Experimental Child Psychology*, *130*, 176–192. https://doi.org/10.1016/j.jecp.2014.10.004

Ho, M. K., Cushman, F., Littman, M. L., & Austerweil, J. L. (2021). Communication in action: Planning and interpreting communicative demonstrations. *Journal of Experimental Psychology. General*, *150*(11), 2246–2272. https://doi.org/10.1037/xge0001035

Holler, J., & Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics*, *43*(14), 3522–3536. https://doi.org/10.1016/j.pragma.2011.08.002

Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research Psychologische Forschung*, *73*(4), 512–526. https://doi.org/10.1007/s00426-009-0234-2

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878. https://doi.org/10.1017/S0140525X01000103

Hudson, M., Bach, P., & Nicholson, T. (2018). You said you would! The predictability of other's behavior from their intentions determines predictive biases in action perception. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(2), 320–335. https://doi.org/10.1037/xhp0000451

Hudson, M., McDonough, K. L., Edwards, R., & Bach, P. (2018). Perceptual teleology: Expectations of action efficiency bias social perception. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1884), 20180638. https://doi.org/10.1098/rspb.2018.0638

Jara-Ettinger, J., & Rubio-Fernandez, P. (2022). The social basis of referential communication: Speakers construct physical reference based on listeners' expected visual search. *Psychological Review*, *129*(6), 1394–1413. https://doi.org/10.1037/rev0000345

Jeannerod, M. (1984). The Timing of Natural Prehension Movements. *Journal of Motor Behavior*, *16*(3), 235–254. https://doi.org/10.1080/00222895.1984.10735319

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*(2), 187–202. https://doi.org/10.1017/S0140525X00034026

Jeannerod, M. (1997). *The cognitive neuroscience of action* (pp. xiv, 236). Blackwell Publishing.

Johnson-Frey, S., McCarty, M., & Keen, R. (2004). Reaching beyond spatial perception: Effects of intended future actions on visually guided prehension. *Visual Cognition*, *11*(2–3), 371–399. https://doi.org/10.1080/13506280344000329

Keller, R. (1998). *A Theory of Linguistic Signs*. Oxford University Press.

Kendon, A. (1991). Some Considerations for a Theory of Language Origins. *Man*, *26*(2), 199. https://doi.org/10.2307/2803829

Kendrick, K. H., Brown, P., Dingemanse, M., Floyd, S., Gipper, S., Hayano, K., Hoey, E., Hoymann, G., Manrique, E., Rossi, G., & Levinson, S. C. (2020). Sequence organization: A universal infrastructure for social action. *Journal of Pragmatics*, *168*, 119–138. https://doi.org/10.1016/j.pragma.2020.06.009

Kendrick, K. H., Holler, J., & Levinson, S. C. (2023). Turn-taking in human face-to-face interaction is multimodal: Gaze direction and manual gestures aid the coordination of turn transitions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1875), 20210473. https://doi.org/10.1098/rstb.2021.0473

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S.-J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, *7*(12), 1299–1301. https://doi.org/10.1038/nn1355

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245–266. https://doi.org/10.1037/rev0000059

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological Research on Joint Action. In *Psychology of Learning and Motivation* (Vol. 54, pp. 59–101). Elsevier. https://doi.org/10.1016/B978-0-12-385527-5.00003-6

Knoblich, G., & Sebanz, N. (2006). The Social Nature of Perception and Action. *Current Directions in Psychological Science*, *15*(3), 99–104. https://doi.org/10.1111/j.0963-7214.2006.00415.x

Koul, A., Cavallo, A., Ansuini, C., & Becchio, C. (2016). Doing It Your Way: How Individual Movement Styles Affect Action Prediction. *PLOS ONE*, *11*(10), e0165297. https://doi.org/10.1371/journal.pone.0165297

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*(5), 113–114. https://doi.org/10.3758/BF03342817

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, *4*(1). https://doi.org/10.1177/2515245920951503

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT Press.

Levinson, S. C. (2016). Turn-taking in Human Communication – Origins and Implications for Language Processing. *Trends in Cognitive Sciences*, *20*(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Levinson, S. C., & Enfield, N. J. (Eds.). (2020). *Roots of Human Sociality: Culture, Cognition and Interaction*. Routledge. https://doi.org/10.4324/9781003135517

Levinson, S. C., & Torreira, F. (2015). Timing in Turn-Taking and its Implications for Processing Models of Language. *Frontiers in Psychology*, *6*, 136034. https://doi.org/10.3389/fpsyg.2015.00731

Lewkowicz, D., Delevoye-Turrell, Y., Bailly, D., Andry, P., & Gaussier, P. (2013). Reading motor intention through mental imagery. *Adaptive Behavior*, *21*(5), 315–327. https://doi.org/10.1177/1059712313501347

Lewkowicz, D., & Delevoye-Turrell, Y. N. (2020). Predictable real-time constraints reveal anticipatory strategies of coupled planning in a sequential pick and place task. *Quarterly Journal of Experimental Psychology*, *73*(4), 594–616. https://doi.org/10.1177/1747021819888081

Lewkowicz, D., Quesque, F., Coello, Y., & Delevoye-Turrell, Y. N. (2015). Individual differences in reading social intentions from motor deviants. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01175

Magyari, L., & de Ruiter, J. P. (2012). Prediction of Turn-Ends Based on Anticipation of Upcoming Words. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00376

Malinowski, B. (1949). *The Problem of Meaning in Primitive Languages: Supplement 1*. Routledge & Kegan Paul Limited.

Manera, V., Becchio, C., Cavallo, A., Sartori, L., & Castiello, U. (2011). Cooperation or competition? Discriminating between social intentions by observing prehensile movements. *Experimental Brain Research*, *211*(3–4), 547–556. https://doi.org/10.1007/s00221-011-2649-4

Manera, V., Becchio, C., Schouten, B., Bara, B. G., & Verfaillie, K. (2011). Communicative Interactions Improve Visual Detection of Biological Motion. *PLoS ONE, 6*(1), e14594. https://doi.org/10.1371/journal.pone.0014594

Manera, V., Schouten, B., Verfaillie, K., & Becchio, C. (2013). Time Will Show: Real Time Predictions during Interpersonal Action Perception. *PLoS ONE*, *8*(1), e54949. https://doi.org/10.1371/journal.pone.0054949

Marteniuk, R. G., Mackenzie, C. L., Jeannerod, M., Athenes, S., & Dugas, C. (1987). Constraints on human arm movement trajectories. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *41*(3), 365–378. https://doi.org/10.1037/h0084157

Maselli, A., Dhawan, A., Cesqui, B., Russo, M., Lacquaniti, F., & d'Avella, A. (2017). Where Are You Throwing the Ball? I Better Watch Your Body, Not Just Your Arm! *Frontiers in Human Neuroscience, 11*, 505. https://doi.org/10.3389/fnhum.2017.00505

McDonough, K. L., Hudson, M., & Bach, P. (2019). Cues to intention bias action perception toward the most efficient trajectory. *Scientific Reports*, *9*(1), 6472. https://doi.org/10.1038/s41598-019-42204-y

McEllin, L., Knoblich, G., & Sebanz, N. (2018). Distinct kinematic markers of demonstration and joint action coordination? Evidence from virtual xylophone playing. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(6), 885–897. https://doi.org/10.1037/xhp0000505

McEllin, L., Sebanz, N., & Knoblich, G. (2018). Identifying others' informative intentions from movement kinematics. *Cognition*, *180*, 246–258. https://doi.org/10.1016/j.cognition.2018.08.001

Melnyk, A. A., Borysenko, V. Ph., & Hénaff, P. (2014). Analysis of synchrony of a handshake between humans. *2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 1753–1758. https://doi.org/10.1109/AIM.2014.6878337

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology: CB, 20*(8), 750–756. https://doi.org/10.1016/j.cub.2010.02.045

Nölle, J., & Galantucci, B. (2022). Experimental semiotics: Past, present, and future. In *The Routledge Handbook of Semiosis and the Brain*. Routledge.

Nota, N., Trujillo, J. P., & Holler, J. (2021). Facial Signals and Social Actions in Multimodal Face-to-Face Interaction. *Brain Sciences*, *11*(8), Article 8. https://doi.org/10.3390/brainsci11081017

Novack, M. A., & Goldin-Meadow, S. (2017). Gesture as representational action: A paper about function. *Psychonomic Bulletin & Review*, *24*(3), 652–665. https://doi.org/10.3758/s13423-016-1145-z

Novack, M. A., Wakefield, E. M., & Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition*, *146*, 339–348.

Noveck, I. (2018). *Experimental Pragmatics: The Making of a Cognitive Science*. Cambridge University Press. https://doi.org/10.1017/9781316027073

Ostrand, R., & Chodroff, E. (2021). It’s alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. *Journal of Phonetics*, *88*, 101074. https://doi.org/10.1016/j.wocn.2021.101074

Pacherie, E., & Dokic, J. (2006). From mirror neurons to joint actions. *Cognitive Systems Research*, *7*(2), 101–112. https://doi.org/10.1016/j.cogsys.2005.11.012

Peirce, C. S. (1955). *Philosophical Writings of Peirce*. Courier Corporation.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pezzulo, G., & Dindo, H. (2011). What should I do next? Using shared representations to solve interaction problems. *Experimental Brain Research*, *211*(3–4), 613–630. https://doi.org/10.1007/s00221-011-2712-1

Pezzulo, G., Donnarumma, F., & Dindo, H. (2013). Human Sensorimotor Communication: A Theory of Signaling in Online Social Interactions. *PLoS ONE*, *8*(11), e79876. https://doi.org/10.1371/journal.pone.0079876

Pezzulo, G., Donnarumma, F., Dindo, H., D’Ausilio, A., Konvalinka, I., & Castelfranchi, C. (2019). The body talks: Sensorimotor communication and its brain and kinematic signatures. *Physics of Life Reviews*, *28*, 1–21. https://doi.org/10.1016/j.plrev.2018.06.014

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154. https://doi.org/10.1080/713752551

Prinz, W., & Rapinett, G. (2008). Filling the gap: Dynamic representation of occluded action. In *Enacting intersubjectivity: A cognitive and social perspective on the study of interactions* (pp. 223–236). IOS Press.

Quesque, F., Delevoye-Turrell, Y., & Coello, Y. (2016). Facilitation effect of observed motor deviants in a cooperative motor task: Evidence for direct perception of social intention in action. *Quarterly Journal of Experimental Psychology*, *69*(8), 1451–1463. https://doi.org/10.1080/17470218.2015.1083596

Quesque, F., Lewkowicz, D., Delevoye-Turrell, Y. N., & Coello, Y. (2013). Effects of social intention on movement kinematics in cooperative actions. *Frontiers in Neurorobotics*, *7*. https://doi.org/10.3389/fnbot.2013.00014

R Core Team, R. (2013). *R: A language and environment for statistical computing*.

Rand, M. K., Alberts, J. L., Stelmach, G. E., & Bloedel, J. R. (1997). The influence of movement segment difficulty on movements with two-stroke sequence: *Experimental Brain Research*, *115*(1), 137–146. https://doi.org/10.1007/PL00005673

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science*, *44*(11), e12911. https://doi.org/10.1111/cogs.12911

Rasenberg, M., Pouw, W., Özyürek, A., & Dingemanse, M. (2022). The multimodal nature of communicative efficiency in social interaction. *Scientific Reports*, *12*(1), 19111. https://doi.org/10.1038/s41598-022-22883-w

Riest, C., Jorschick, A. B., & de Ruiter, J. P. (2015). Anticipation in turn-taking: Mechanisms and information sources. *Frontiers in Psychology*, *6*, 89. https://doi.org/10.3389/fpsyg.2015.00089

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188–194. https://doi.org/10.1016/S0166-2236(98)01260-0

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews. Neuroscience*, *11*(4), 264–274. https://doi.org/10.1038/nrn2805

Roepstorff, A., & Frith, C. (2004). What?s at the top in the top-down control of action? Script-sharing and ?top-top? Control of action in cognitive experiments. *Psychological Research*, *68*(2–3), 189–198. https://doi.org/10.1007/s00426-003-0155-4

Rosenbaum, D. A., Marchak, F., Barnes, H. J., Vaughan, J., Slotta, J. D., & Jorgensen, M. J. (1990). Constraints for action selection: Overhand versus underhand grips. In *Attention* *and performance 13: Motor representation and control* (pp. 321–342). Lawrence Erlbaum Associates, Inc.

Rosenbaum, D. A., Vaughan, J., Barnes, H. J., & Jorgensen, M. J. (1992). Time course of movement planning: Selection of handgrips for object manipulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1058–1073. https://doi.org/10.1037/0278-7393.18.5.1058

Rosenbaum, D., Marchak, F., Barnes, H., Vaughan, J., Slotta, J., & Jorgensen, M. (1990). Constraints for action selection: Overhand versus underhand grips. *Attention and Performance XIII, Vol. XIII*.

Royka, A., Chen, A., Aboody, R., Huanca, T., & Jara-Ettinger, J. (2022). People infer communicative action through an expectation for efficient communication. *Nature Communications*, *13*(1), 4160. https://doi.org/10.1038/s41467-022-31716-3

Sacheli, L. M., Tidoni, E., Pavone, E. F., Aglioti, S. M., & Candidi, M. (2013). Kinematics fingerprints of leader and follower role-taking during cooperative joint actions. *Experimental Brain Research*, *226*(4), 473–486. https://doi.org/10.1007/s00221-013-3459-7

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A Simple Systematic for the Organisation of Turn Taking in Conversation. *Language*, *50*, 696–735. https://doi.org/10.2307/412243

Santamaria, J. P., & Rosenbaum, D. A. (2011). Etiquette and effort: Holding doors for others. *Psychological Science*, *22*(5), 584–588. https://doi.org/10.1177/0956797611406444

Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: The role of movement information. *Cognition*, *119*(2), 242–252. https://doi.org/10.1016/j.cognition.2011.01.014

Schegloff, E. (1982). Discourse as an interactional achievement: Some uses of `uh huh' and other things that come between sentences. In *Analyzing discourse: Text and talk, Georgetown University Roundtable on Languages and Linguistics* (pp. 71–93).

Schegloff, E. A. (1991). Conversation analysis and socially shared cognition. In *Perspectives on socially shared cognition* (pp. 150–171). American Psychological Association. https://doi.org/10.1037/10096-007

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis, Vol. 1* (pp. xv, 300). Cambridge University Press. https://doi.org/10.1017/CBO9780511791208

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *The Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/S0140525X12000660

Schmitz, L. (2017). *Co-representation and Communication in Joint Action* [PhD Thesis, Doctoral dissertation]. Budapest: Central European University]. https://www.etd.ceu.edu/2019/schmitz_laura.pdf

Schmitz, L., Knoblich, G., Deroy, O., & Vesper, C. (2021). Crossmodal correspondences as common ground for joint action. *Acta Psychologica*, *212*, 103222. https://doi.org/10.1016/j.actpsy.2020.103222

Schmitz, L., Vesper, C., Sebanz, N., & Knoblich, G. (2018a). Co-actors represent the order of each other's actions. *Cognition*, *181*, 65–79. https://doi.org/10.1016/j.cognition.2018.08.008

Schmitz, L., Vesper, C., Sebanz, N., & Knoblich, G. (2018b). When Height Carries Weight: Communicating Hidden Object Properties for Joint Action. *Cognitive Science*, *42*(6), 2021–2059. https://doi.org/10.1111/cogs.12638

Schubotz, R. I., Wurm, M. F., Wittmann, M. K., & von Cramon, D. Y. (2014). Objects tell us what action we can expect: Dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00636

Schultz, B. G., O'brien, I., Phillips, N., McFARLAND, D. H., Titone, D., & Palmer, C. (2016). Speech rates converge in scripted turn-taking conversations. *Applied Psycholinguistics*, *37*(5), 1201–1220. https://doi.org/10.1017/S0142716415000545

Scott-Phillips, T. (2014). *Speaking Our Minds: Why human communication is different, and how language evolved to make it special*. Bloomsbury Publishing.

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. S. (2009). Signalling signalhood and the emergence of communication. *Cognition*, *113*(2), 226–233. https://doi.org/10.1016/j.cognition.2009.08.009

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Sebanz, N., & Knoblich, G. (2009). Prediction in Joint Action: What, When, and Where. *Topics in Cognitive Science*, *1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x

Sebanz, N., & Knoblich, G. (2021). Progress in Joint-Action Research. *Current Directions in Psychological Science*, *30*(2), 138–143. https://doi.org/10.1177/0963721420984425

Sinigaglia, C., & Butterfill, S. (2016). Motor representation in goal ascription. In *Foundations of embodied cognition: Conceptual and interactive embodiment* (pp. 149–164). Routledge/Taylor & Francis Group.

Sinigaglia, C., & Butterfill, S. A. (2022). Motor representation in acting together. *Synthese*, *200*(2), 82. https://doi.org/10.1007/s11229-022-03539-8

Sparenberg, P., Springer, A., & Prinz, W. (2012). Predicting others' actions: Evidence for a constant time delay in action simulation. *Psychological Research*, *76*(1), 41–49. https://doi.org/10.1007/s00426-011-0321-z

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition, 2nd ed* (pp. viii, 326). Blackwell Publishing.

Sperber, D., & Wilson, D. (2015). Beyond Speaker?s Meaning. *Croatian Journal of Philosophy*, *15*(2), 117–149.

Springer, A., Brandstädter, S., & Prinz, W. (2013). Dynamic Simulation and Static Matching for Action Prediction: Evidence From Body Part Priming. *Cognitive Science*, *37*(5), 936–952. https://doi.org/10.1111/cogs.12044

Springer, A., Parkinson, J., & Prinz, W. (2013). Action simulation: Time course and representational mechanisms. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00387

Springer, A., & Prinz, W. (2010). Action Semantics Modulate Action Prediction. *Quarterly Journal of Experimental Psychology*, *63*(11), 2141–2158. https://doi.org/10.1080/17470211003721659

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

Taub, S. F. (2001). *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge University Press.

Thornton, I. M., & Knoblich, G. (2006). Action Perception: Seeing the world through a moving body. *Current Biology: CB*, *16*(1), R27-29. https://doi.org/10.1016/j.cub.2005.12.006

Thornton, M. A., & Tamir, D. I. (2021). People accurately predict the transition probabilities between actions. *Science Advances*, *7*(9), eabd4995. https://doi.org/10.1126/sciadv.abd4995

Tomasello, M. (2008). *Origins of human communication* (pp. xiii, 393). MIT Press.

Török, G., Pomiechowska, B., Csibra, G., & Sebanz, N. (2019). Rationality in Joint Action: Maximizing Coefficiency in Coordination. *Psychological Science*, *30*(6), 930–941. https://doi.org/10.1177/0956797619842550

Trujillo, J. P., & Holler, J. (2023). Interactionally Embedded Gestalt Principles of Multimodal Human Communication. *Perspectives on Psychological Science*, *18*(5), 1136–1159. https://doi.org/10.1177/17456916221141422

Trujillo, J. P., Simanova, I., Bekkering, H., & Özyürek, A. (2018). Communicative intent modulates production and comprehension of actions and gestures: A Kinect study. *Cognition*, *180*, 38–51. https://doi.org/10.1016/j.cognition.2018.04.003

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., Hristova, D., Karlinsky, A., McEllin, L., Nijssen, S. R. R., Schmitz, L., & Wahn, B. (2017). Joint Action: Mental Representations, Shared Information and General Mechanisms for Coordinating with Others. *Frontiers in Psychology*, *07*. https://doi.org/10.3389/fpsyg.2016.02039

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*(8–9), 998–1003. https://doi.org/10.1016/j.neunet.2010.06.002

Vesper, C., Morisseau, T., Knoblich, G., & Sperber, D. (2021). When is ostensive communication used for joint action? *Cognitive Semiotics*, *14*(2), 101–129. https://doi.org/10.1515/cogsem-2021-2040

Vesper, C., & Richardson, M. J. (2014). Strategic communication and behavioral coupling in asymmetric joint action. *Experimental Brain Research*, *232*(9), 2945–2956. https://doi.org/10.1007/s00221-014-3982-1

Vesper, C., Schmitz, L., & Knoblich, G. (2017). Modulating action duration to establish nonconventional communication. *Journal of Experimental Psychology: General*, *146*(12), 1722–1737. https://doi.org/10.1037/xge0000379

Vesper, C., Schmitz, L., Knoblich, G., & Sebanz, N. (2016, July 1). *Using Violations of Fitts' Law to Communicate during Joint Action*.

Vesper, C., Schmitz, L., Safra, L., Sebanz, N., & Knoblich, G. (2016). The role of shared visual information for joint action coordination. *Cognition*, *153*, 118–123. https://doi.org/10.1016/j.cognition.2016.05.002

Vesper, C., & Sevdalis, V. (2020). Informing, Coordinating, and Performing: A Perspective on Functions of Sensorimotor Communication. *Frontiers in Human Neuroscience*, *14*, 168. https://doi.org/10.3389/fnhum.2020.00168

Vesper, C., Soutschek, A., & Schubö, A. (2009). Motion coordination affects movement parameters in a joint pick-and-place task. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470210902919067

Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, *211*(3–4), 517–530. https://doi.org/10.1007/s00221-011-2706-z

Wharton, T. (2009). *Pragmatics and Non-Verbal Communication*. Cambridge University Press.

Wilbur, R. B. (1987). *American Sign Language: Linguistic and applied dimensions*, *2nd ed* (pp. xii, 387). Little, Brown and Co.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*(2), 183–194. https://doi.org/10.1016/0749-596X(92)90010-U

Wilson, M., & Knoblich, G. (2005). The Case for Motor Involvement in Perceiving Conspecifics. *Psychological Bulletin*, *131*(3), 460–473. https://doi.org/10.1037/0033-2909.131.3.460

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 593–602. https://doi.org/10.1098/rstb.2002.1238

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), R729–R732. https://doi.org/10.1016/S0960-9822(01)00432-8

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7–8), 1317–1329. https://doi.org/10.1016/S0893-6080(98)00066-5